(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46209 A2

(51) International Patent Classification[7]: C07K

(21) International Application Number: PCT/US01/47218

(22) International Filing Date: 7 December 2001 (07.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/254,367 8 December 2000 (08.12.2000) US
60/288,470 3 May 2001 (03.05.2001) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ANASTASIO, Alison E.** [US/US]; 367 Alden Avenue #8B, New Haven, CT 06515 (US). **HAN, Jin-Hua** [CN/US]; 569B Prospect Street, New Haven, CT 06511 (US). **KLIEM, Stefanie E.** [DE/DE]; Kiefernweg 37, 61440 Oberusel (DE). **ROUNDS, Eileen** [US/US]; 36 Patton Drive, Cheshire, CT 06410 (US).

(74) Agents: **FIELD, Gisela M.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46209 A2

(54) Title: HAPLOTYPES OF THE CYP3A5 GENE

(57) Abstract: Novel genetic variants of the Cytochrome P450, Subfamily IIIA, Polypeptide 5 (CYP3A5) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the CYP3A5 gene. Compositions and methods for haplotyping and/or genotyping the CYP3A5 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE CYP3A5 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/288,470 filed May 3, 2001 and U.S. Provisional Application Serial No. 60/254,367 filed December 8, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of

such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene involved in the metabolism of drugs is the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene or its encoded product. CYP3A5 is an enzyme that belongs to the cytochrome P450 family, a group of heme-thiolate monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids, lipids and xenobiotics. CYP3A enzymes are involved in an NADPH-dependent electron transport pathway, are

the most abundantly expressed cytochrome P450 enzymes in the liver, and are responsible for the metabolism of over 50% of all clinically used drugs (Paulussen et al., *Pharmacogenetics* 2000, 10(5):415-24 2000). CYP3A5 localizes to the endoplasmic reticulum and its expression is induced by glucocorticoids and some pharmacological agents (NCBI Locus Link: Locus ID#1577). The expression and activity of CYP3A5 shows wide interindividual variation, influencing both drug response and disease susceptibility.

By screening a liver cDNA library with CYP3A4 as probe, Aoyama et al. (*J. Biol. Chem.* 264: 10388-10395, 1989) isolated a cDNA encoding CYP3A5. Immunoblot analysis of liver microsomes showed that CYP3A5 is expressed as a 52.5-kD protein, whereas CYP3A4 migrates as a 52.0-kD protein. The CYP3A5 protein was shown to share an 85% sequence similarity with CYP3A4. Analysis of enzymatic activity revealed that CYP3A4 and CYP3A5 have overlapping substrate specificity with minor differences in the metabolism of steroids and drug substrates.

The cytochrome P450, subfamily IIIA, polypeptide 5 gene is located on chromosome 7q21.1 and contains 13 exons that encode a 502 amino acid protein. A reference sequence for the CYP3A5 gene is shown in the contiguous lines of Figure 1 (Genaissance Reference No. 1225874; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_000777.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Several polymorphisms of the CYP3A5 gene have been previously identified. These single nucleotide polymorphisms correspond to the sites named PS3, PS4, PS15, and PS25 herein. Specifically, the variation which corresponds to PS3 consists of a guanine or adenine at nucleotide position 3927 in Figure 1 (Kuehl et al., *Nat Genet* 2001, 27(4):383-91). The presence of the CYP3A5*1C allele, which corresponds to PS4, consists of a cytosine or thymine at nucleotide position 3939 in Figure 1 and is associated with high levels of active CYP3A5 (Kuehl et al., *supra*). Kuehl et al. (*supra*) also demonstrated that polymorphisms in the CYP3A5 gene, designated CYP3A5*3 and CYP3A5*6, result in splice variants and protein truncation. The CYP3A5*6 allele corresponds to PS15 and consists of a guanine or adenine at nucleotide position 18697 in Figure 1. The variation which corresponds to PS25 consists of a thymine or cytosine at nucleotide position 35618 in Figure 1 (NCBI SNP ID: rs15524). As a result of the CYP3A5*3 and CYP3A5*6 polymorphisms, CYP3A5 fails to accumulate in tissues of some people. All Caucasian individuals and most African Americans homozygous (-\-) for CYP3A5*3 had CYP3A5 protein levels less than 21 pmol/mg of protein. However, the presence of at least one CYP3A5*1 allele resulted in CYP3A5 levels ranging from 21-202 pmol/mg of protein (Kuehl et al., *supra*). The polymorphic distribution of the CYP3A5*1 allele indicates that relatively high levels of metabolically active CYP3A5 are expressed by an estimated 30% of Caucasians, 30% of Japanese, 30% of Mexicans, 40% of Chinese, and more than 50% of African Americans, Pacific Islanders, Southeast Asians, and Southwestern American Indians. Since CYP3A5 represents 50% of total hepatic CYP3A content, it may be may be the most important

genetic contributor to interindividual and interracial differences in CYP3A-dependent drug clearance and in responses to many medicines (Kuehl et al., *supra*).

Because of the potential for variation in the CYP3A5 gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the CYP3A5 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of CYP3A5 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 21 novel polymorphic sites in the CYP3A5 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 3633 (PS1), 3747 (PS2), 3998 (PS5), 7657 (PS6), 7717 (PS7), 7830 (PS8), 9523 (PS9), 11189 (PS10), 11214 (PS11), 11310 (PS12), 16830 (PS13), 17383 (PS14), 18727 (PS16), 18787 (PS17), 19755 (PS18), 19806 (PS19), 20065 (PS20), 21170 (PS21), 31057 (PS22), 33640 (PS23) and 35506 (PS24). The polymorphisms at these sites are adenine or guanine at PS1, cytosine or guanine at PS2, adenine or cytosine at PS5, thymine or cytosine at PS6, cytosine or thymine at PS7, guanine or adenine at PS8, thymine or adenine at PS9, cytosine or adenine at PS10, cytosine or thymine at PS11, cytosine or adenine at PS12, cytosine or thymine at PS13, guanine or adenine at PS14, adenine or guanine at PS16, cytosine or thymine at PS17, cytosine or thymine at PS18, thymine or cytosine at PS19, adenine or cytosine at PS20, guanine or thymine at PS21, adenine or guanine at PS22, guanine or adenine at PS23 and thymine or cytosine at PS24. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified sites at nucleotide positions 3927 (PS3), 3939 (PS4), 18697 (PS15) and 35618 (PS25), in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS25 in the CYP3A5 gene, which are shown below in Tables 5 and 4, respectively. Each of these CYP3A5 haplotypes constitutes a code that defines the variant nucleotides that exist in the human population at this set of polymorphic sites in the CYP3A5 gene. Thus each CYP3A5 haplotype also represents a naturally-occurring isoform (also referred to herein as an "isogene") of the CYP3A5 gene. The frequency of each haplotype and haplotype pair within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the CYP3A5 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20,

PS21, PS22, PS23 and PS24 in both copies of the CYP3A5 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel CYP3A5 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel CYP3A5 polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-PS25. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the CYP3A5 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the CYP3A5 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's CYP3A5 gene is defined by one of the CYP3A5 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's CYP3A5 gene are defined by one of the CYP3A5 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. Establishing the CYP3A5 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with CYP3A5 activity, e.g., drug metabolizing disorders.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate CYP3A5 as a candidate target for treating a specific condition or disease predicted to be associated with CYP3A5 activity. Determining for a particular population the frequency of one or more of the individual CYP3A5 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue CYP3A5 as a target for treating the specific disease of interest. In particular, if variable CYP3A5 activity is associated with the disease, then one or more CYP3A5 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed CYP3A5 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable CYP3A5 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any CYP3A5 haplotype or haplotype pair, apply the information derived from detecting CYP3A5 haplotypes in an individual to decide whether modulating CYP3A5 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting CYP3A5 to treat a specific condition or disease predicted to be associated with CYP3A5 activity. For example, detecting which of the CYP3A5 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to

screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the CYP3A5 isoforms present in the disease population, or for only the most frequent CYP3A5 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular CYP3A5 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

Haplotyping the CYP3A5 gene in an individual is also useful to control for genetically-based bias in the design of candidate drugs that target or are metabolized by CYP3A5. For example, for a lead compound that is metabolized by CYP3A5, the pharmaceutical scientist of ordinary skill would be concerned that a favorable efficacy and/or side effect profile shown in a Phase II or Phase III trial may not be replicated in the general population if a higher (or lower) percentage of patients in the treatment group, compared to the general population, have a form of the CYP3A5 gene that makes them genetically predisposed to metabolize the drug more efficiently than patients with other forms of the CYP3A5 gene. Similarly, this pharmaceutical scientist would recognize the potential for bias in the results of a Phase II or Phase III clinical trial of a drug targeting CYP3A5 that could be introduced if individuals whose CYP3A5 gene structure makes them genetically predisposed to respond well to the drug are present in a higher (or lower) frequency in the treatment group than in the control group (Bacanu et al., 2000, *Am. J. Hum. Gen.* 66:1933-44; Pritchard et al., 2000, *Am. J. Hum. Gen.* 67: 170-81).

The pharmaceutical scientist can immediately reduce this potential for genetically-base bias in the results of clinical trials of drugs metabolized by or targeting CYP3A5 by practicing the claimed invention. In particular, by determining which of the CYP3A5 haplotypes disclosed herein are present in individuals recruited to participate in a clinical trial of a drug metabolized by or targeting CYP3A5, the pharmaceutical scientist can then assign that individual to the treatment or control group as appropriate to ensure that approximately equal frequencies of different CYP3A5 haplotypes (or haplotype pairs) are represented in the two groups and/or the frequencies of different CYP3A5 haplotypes or haplotype pairs are similar to the frequencies in the general population. Thus, by practicing the claimed invention, the pharmaceutical scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any CYP3A5 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a CYP3A5 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the CYP3A5 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the CYP3A5 genotype or haplotype in a reference population. A higher frequency of the CYP3A5 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the CYP3A5 genotype, haplotype, or haplotype pair. In preferred

embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the CYP3A5 haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for drug metabolizing disorders.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the CYP3A5 gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, guanine at PS2, cytosine at PS5, cytosine at PS6, thymine at PS7, adenine at PS8, adenine at PS9, adenine at PS10, thymine at PS11, adenine at PS12, thymine at PS13, adenine at PS14, guanine at PS16, thymine at PS17, thymine at PS18, cytosine at PS19, cytosine at PS20, thymine at PS21, guanine at PS22, adenine at PS23 and cytosine at PS24. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of adenine at PS3, thymine at PS4, adenine at PS15 and cytosine at PS25.

A particularly preferred polymorphic variant is an isogene of the CYP3A5 gene. A CYP3A5 isogene of the invention comprises adenine or guanine at PS1, cytosine or guanine at PS2, guanine or adenine at PS3, cytosine or thymine at PS4, adenine or cytosine at PS5, thymine or cytosine at PS6, cytosine or thymine at PS7, guanine or adenine at PS8, thymine or adenine at PS9, cytosine or adenine at PS10, cytosine or thymine at PS11, cytosine or adenine at PS12, cytosine or thymine at PS13, guanine or adenine at PS14, guanine or adenine at PS15, adenine or guanine at PS16, cytosine or thymine at PS17, cytosine or thymine at PS18, thymine or cytosine at PS19, adenine or cytosine at PS20, guanine or thymine at PS21, adenine or guanine at PS22, guanine or adenine at PS23, thymine or cytosine at PS24 and thymine or cytosine at PS25. The invention also provides a collection of CYP3A5 isogenes, referred to herein as a CYP3A5 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a CYP3A5 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 88, adenine at a position corresponding to nucleotide 299 and guanine at a position corresponding to nucleotide 654. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of adenine at a position corresponding to nucleotide 624. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a CYP3A5 isogene defined by haplotypes 2, 5, 7-8, 18-19, and 21.

Polynucleotides complementary to these CYP3A5 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the CYP3A5 gene will be useful in studying the expression and function of CYP3A5, and in expressing CYP3A5 protein for use in screening for candidate drugs to treat diseases related to CYP3A5 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one

of the polymorphic genomic and cDNA variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express CYP3A5 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the CYP3A5 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of tyrosine at a position corresponding to amino acid position 30 and tyrosine at a position corresponding to amino acid position 100. A polymorphic variant of CYP3A5 is useful in studying the effect of the variation on the biological activity of CYP3A5 as well as on the binding affinity of candidate drugs to CYP3A5, or studying the enzymatic properties of such CYP3A5 variants using these candidate drugs as substrates. Herein, the term drug refers to a candidate drug or any of its metabolic derivatives.

The present invention also provides antibodies that recognize and bind to the above polymorphic CYP3A5 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one or more of the CYP3A5 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the CYP3A5 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against CYP3A5 protein, and for testing the efficacy of therapeutic agents and compounds for drug metabolizing disorders in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the CYP3A5 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes one or more of the following: the polymorphisms, the genotypes, the haplotypes, and the haplotype pairs identified for the CYP3A5 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing CYP3A5 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the CYP3A5 gene (Genaissance Reference No. 1225874; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO

standard ST.25). SEQ ID NO:109 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS25, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:109 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the CYP3A5 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the CYP3A5 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the CYP3A5 gene. As described in more detail below, the inventors herein discovered 26 isogenes of the CYP3A5 gene by characterizing the CYP3A5 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The CYP3A5 isogenes present in the human reference population are defined by haplotypes for 25 polymorphic sites in the CYP3A5 gene, 21 of which are believed to be novel. The CYP3A5 polymorphic sites identified by the inventors are referred to as PS1-PS25 to designate the order in which they are located in the gene (see Table 3 below), with the novel polymorphic sites referred to as PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24. Using the genotypes identified in the Index Repository for PS1-PS25 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the CYP3A5 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the CYP3A5 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether CYP3A5 is a suitable target for drugs to treat drug metabolizing disorders, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular

nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA, coding sequence or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms (e.g., alleles) of a gene found in a population. An isogene (or allele) contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to

absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

Polymorphic variant (or **variant**)– A gene, mRNA, cDNA, polypeptide, protein or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or

disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the CYP3A5 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel CYP3A5 polymorphisms, haplotypes and haplotype pairs identified herein.

The compositions comprise at least one oligonucleotide for detecting the variant nucleotide or nucleotide pair located at a novel CYP3A5 polymorphic site in one copy or two copies of the CYP3A5 gene. Such oligonucleotides are referred to herein as CYP3A5 haplotyping oligonucleotides or genotyping oligonucleotides, respectively, and collectively as CYP3A5 oligonucleotides. In one embodiment, a CYP3A5 haplotyping or genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that contains, or that is located close to, one of the novel polymorphic sites described herein.

As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Haplotyping or genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a CYP3A5 polynucleotide. Preferably, the target region is located in a CYP3A5 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with another region in the CYP3A5 polynucleotide or

with a non-CYP3A5 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the CYP3A5 gene using the polymorphism information provided herein in conjunction with the known sequence information for the CYP3A5 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred haplotyping or genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th

position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention. ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains either of the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting CYP3A5 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GCTTGTGRGGATGGA (SEQ ID NO:4) and its complement,
CCAGAACSCTTGGAC (SEQ ID NO:5) and its complement,
CAGTTGAMGAAGGAA (SEQ ID NO:6) and its complement,
TGATCTAYAAAGTCA (SEQ ID NO:7) and its complement,
CCGTACAYATGGACT (SEQ ID NO:8) and its complement,
TCTTATGRTTGCAAA (SEQ ID NO:9) and its complement,
AAGAGGAWAATTACT (SEQ ID NO:10) and its complement,
GCAGAATMGGGCTAG (SEQ ID NO:11) and its complement,
TCAGCTCYGTTGTCC (SEQ ID NO:12) and its complement,
TGTTATTMTGTCTTC (SEQ ID NO:13) and its complement,
AATGTTTYTGTTGAA (SEQ ID NO:14) and its complement,
GACAGTCRCACTGTT (SEQ ID NO:15) and its complement,
TAGATCCRTTATTTC (SEQ ID NO:16) and its complement,
ATAACTGYTTTCTTG (SEQ ID NO:17) and its complement,
ATAATTGYTCCAGGT (SEQ ID NO:18) and its complement,
TTGTTTTYCCCACAG (SEQ ID NO:19) and its complement,
GAACAAGMGAAGCCA (SEQ ID NO:20) and its complement,
GCAGGAAKTATTCCA (SEQ ID NO:21) and its complement,
TACTTCARTAGTACT (SEQ ID NO:22) and its complement,
TTTTTATRTTTCATT (SEQ ID NO:23) and its complement, and
ACTATTGYAGATCCC (SEQ ID NO:24) and its complement.

A preferred ASO primer for detecting CYP3A5 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GGTGTGGCTTGTGRG (SEQ ID NO:25); TTGAAATCCATCCYC (SEQ ID NO:26);
AAGAACCCAGAACSC (SEQ ID NO:27); CGGGGAGTCCAAGSG (SEQ ID NO:28);
AGAACACAGTTGAMG (SEQ ID NO:29); GCCACTTTCCTTCKT (SEQ ID NO:30);
GCCCTCTGATCTAYA (SEQ ID NO:31); GGATTGTGACTTTRT (SEQ ID NO:32);
TGGGACCCGTACAYA (SEQ ID NO:33); TTAAAAAGTCCATRT (SEQ ID NO:34);
TTTGCTTCTTATGRT (SEQ ID NO:35); CTGATGTTTGCAAYC (SEQ ID NO:36);
TGAAAGAAGAGGAWA (SEQ ID NO:37); CTCCCAAGTAATTWT (SEQ ID NO:38);
CCAGCTGCAGAATMG (SEQ ID NO:39); ACTTCACTAGCCCKA (SEQ ID NO:40);
GTTTAATCAGCTCYG (SEQ ID NO:41); GTGTGGGGACAACRG (SEQ ID NO:42);
AAAGAATGTTATTMT (SEQ ID NO:43); ATTTGTGAAGACAKA (SEQ ID NO:44);
AGAAAAAATGTTTYT (SEQ ID NO:45); CTAGAGTTCAACARA (SEQ ID NO:46);
GGAGTCGACAGTCRC (SEQ ID NO:47); TAACCCAACAGTGYG (SEQ ID NO:48);
GTTTCTTAGATCCRT (SEQ ID NO:49); TTGAGAGAAATAAYG (SEQ ID NO:50);
TTAAAAATAACTGYT (SEQ ID NO:51); ATATGTCAAGAAARC (SEQ ID NO:52);
AAAATTATAATTGYT (SEQ ID NO:53); AACTTTACCTGGARC (SEQ ID NO:54);

TTTGTTTTGTTTTYC (SEQ ID NO:55); AGAGTACTGTGGGRA (SEQ ID NO:56);
TGTTTAGAACAAGMG (SEQ ID NO:57); ACCAAATGGCTTCKC (SEQ ID NO:58);
AAATGTGCAGGAAKT (SEQ ID NO:59); TCTTCCTGGAATAMT (SEQ ID NO:60);
TTCTAATACTTCART (SEQ ID NO:61); CCATGCAGTACTAYT (SEQ ID NO:62);
CTGTGGTTTTTATRT (SEQ ID NO:63); ATAGTTAATGAAAYA (SEQ ID NO:64);
TGTTTAACTATTGYA (SEQ ID NO:65); and TTCAAGGGGATCTRC (SEQ ID NO:66).

Other oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting CYP3A5 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GTGGCTTGTG (SEQ ID NO:67); AAATCCATCC(SEQ ID NO:68);
AACCCAGAAC (SEQ ID NO:69); GGAGTCCAAG(SEQ ID NO:70);
ACACAGTTGA (SEQ ID NO:71); ACTTTCCTTC(SEQ ID NO:72);
CTCTGATCTA (SEQ ID NO:73); TTGTGACTTT(SEQ ID NO:74);
GACCCGTACA (SEQ ID NO:75); AAAAGTCCAT(SEQ ID NO:76);
GCTTCTTATG (SEQ ID NO:77); ATGTTTGCAA(SEQ ID NO:78);
AAGAAGAGGA (SEQ ID NO:79); CCAAGTAATT(SEQ ID NO:80);
GCTGCAGAAT (SEQ ID NO:81); TCACTAGCCC(SEQ ID NO:82);
TAATCAGCTC (SEQ ID NO:83); TGGGGACAAC(SEQ ID NO:84);
GAATGTTATT (SEQ ID NO:85); TGTGAAGACA(SEQ ID NO:86);
AAAAATGTTT (SEQ ID NO:87); GAGTTCAACA(SEQ ID NO:88);
GTCGACAGTC (SEQ ID NO:89); CCCAACAGTG(SEQ ID NO:90);
TCTTAGATCC (SEQ ID NO:91); AGAGAAATAA(SEQ ID NO:92);
AAAATAACTG (SEQ ID NO:93); TGTCAAGAAA(SEQ ID NO:94);
ATTATAATTG (SEQ ID NO:95); TTTACCTGGA(SEQ ID NO:96);
GTTTTGTTTT (SEQ ID NO:97); GTACTGTGGG(SEQ ID NO:98);
TTAGAACAAG (SEQ ID NO:99); AAATGGCTTC(SEQ ID NO:100);
TGTGCAGGAA (SEQ ID NO:101); TCCTGGAATA(SEQ ID NO:102);
TAATACTTCA (SEQ ID NO:103); TGCAGTACTA(SEQ ID NO:104);
TGGTTTTTAT (SEQ ID NO:105); GTTAATGAAA(SEQ ID NO:106);
TTAACTATTG (SEQ ID NO:107); and AAGGGGATCT(SEQ ID NO:108).

In some embodiments, a composition contains two or more differently labeled CYP3A5 oligonucleotides for simultaneously probing the identity of nucleotides or nucleotide pairs at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

CYP3A5 oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of polymorphism detection assays,

including but not limited to probe hybridization and polymerase extension assays. Immobilized CYP3A5 oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two CYP3A5 oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the CYP3A5 gene in an individual. As used herein, the terms "CYP3A5 genotype" and "CYP3A5 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the CYP3A5 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of a genotyping method of the invention involves isolating from the individual a nucleic acid sample comprising the two copies of the CYP3A5 gene, mRNA transcripts thereof or cDNA copies thereof, or a fragment of any of the foregoing, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24 in the two copies to assign a CYP3A5 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene, mRNA or cDNA (or fragment of such CYP3A5 molecules) in an individual may be the same allele or may be different alleles. In a preferred embodiment of the method for assigning a CYP3A5 genotype, the identity of the nucleotide pair at one or more of the polymorphic sites selected from the group consisting of PS3, PS4, PS15 and PS25 is also determined. In another embodiment, a genotyping method of the invention comprises determining the identity of the nucleotide pair at each of PS1-PS25.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the CYP3A5 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions if not present in the mRNA or cDNA. If a CYP3A5 gene fragment is isolated, it must contain the polymorphic site(s) to be

genotyped.

One embodiment of a haplotyping method of the invention comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the CYP3A5 gene, mRNA or cDNA, or a fragment of such CYP3A5 molecules, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24 in that copy to assign a CYP3A5 haplotype to the individual.

The nucleic acid used in the above haplotyping methods of the invention may be isolated using any method capable of separating the two copies of the CYP3A5 gene or fragment such as one of the methods described above for preparing CYP3A5 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will typically only provide haplotype information on one of the two CYP3A5 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional CYP3A5 clones will usually need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the CYP3A5 gene in an individual. In some cases, however, once the haplotype for one CYP3A5 allele is directly determined, the haplotype for the other allele may be inferred if the individual has a known genotype for the polymorphic sites of interest or if the haplotype frequency or haplotype pair frequency for the individual's population group is known. In some embodiments, the CYP3A5 haplotype is assigned to the individual by also identifying the nucleotide at one or more polymorphic sites selected from the group consisting of PS3, PS4, PS15 and PS25. In a particularly preferred embodiment, the nucleotide at each of PS1-PS25 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the CYP3A5 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's CYP3A5 gene, the phased sequence of nucleotides present at each of PS1-PS25. This identifying step does not necessarily require that each of PS1-PS25 be directly examined. Typically only a subset of PS1-PS25 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In another embodiment of a haplotyping method of the invention, a CYP3A5 haplotype pair is

determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24 in each copy of the CYP3A5 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS25 in each copy of the CYP3A5 gene.

When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the CYP3A5 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988). Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific

oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the CYP3A5 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, mRNA, cDNA or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a

polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by haplotyping or genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Detection of the allele(s) present at a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's CYP3A5 haplotype pair is predicted from its CYP3A5 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a CYP3A5 genotype for the individual at two or more CYP3A5 polymorphic sites described herein, accessing data containing CYP3A5 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the genotype data. In one embodiment, the reference haplotype pairs include the CYP3A5 haplotype pairs shown in Table 4. The CYP3A5 haplotype pair can be assigned by comparing the individual's genotype with the genotypes corresponding to the haplotype pairs known to exist in the general population or in a specific population group, and determining which haplotype pair is consistent with the genotype of the individual. In some embodiments, the comparing step may be performed by visual inspection (for example, by consulting Table 4). When the genotype of the individual is consistent with more than one haplotype pair, frequency data (such as that presented in Table 7) may be used to determine which of these haplotype pairs is most likely to be present in the individual. This determination may also be performed in some embodiments by visual inspection, for example by consulting Table 7. If a particular CYP3A5 haplotype pair consistent with the genotype of the individual is more frequent in the reference population than others consistent with the genotype, then that haplotype pair with the highest frequency is the most likely to be present in the individual. In other embodiments, the comparison may be made by a computer-implemented algorithm with the genotype of the individual and the reference haplotype data stored in computer-readable formats. For example, as described in PCT/US01/12831, filed April 18, 2001, one computer-implemented algorithm to perform this comparison entails enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing CYP3A5 haplotype pairs frequency data determined in a reference

population to determine a probability that the individual has a possible haplotype pair, and analyzing the determined probabilities to assign a haplotype pair to the individual.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a CYP3A5 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the

known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22; copending PCT/US01/12831 filed April 18, 2001) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a CYP3A5 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel CYP3A5 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be e.g., a reference population, a family population, a same gender population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for CYP3A5 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a CYP3A5 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. In one embodiment, the method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one or more of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular CYP3A5 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that CYP3A5 genotype, haplotype or haplotype pair. Preferably, the CYP3A5 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes

and sub-haplotypes derived from these genotypes and haplotypes. Sub-genotypes useful in the invention preferably do not include sub-genotypes solely for any one of PS3, PS4, PS15 and PS25 or for any combination thereof.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting CYP3A5 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and/or adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a CYP3A5 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the CYP3A5 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between

individual response and CYP3A5 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their CYP3A5 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the CYP3A5 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in WO 01/01218, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between CYP3A5 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in WO 01/01218.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the CYP3A5 gene. As described in WO 01/01218, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of CYP3A5 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the CYP3A5 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the CYP3A5 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the

diagnostic test results and the underlying CYP3A5 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the CYP3A5 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant CYP3A5 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, and may also comprise one or more additional polymorphisms selected from the group consisting of adenine at PS3, thymine at PS4, adenine at PS15 and cytosine at PS25. Similarly, the nucleotide sequence of a variant fragment of the CYP3A5 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the CYP3A5 gene, which is defined by haplotype 12, (or other reported CYP3A5 sequences) or to portions of the reference sequence (or other reported CYP3A5 sequences), except for the haplotyping and genotyping oligonucleotides described above.

The location of a polymorphism in a variant CYP3A5 gene or fragment is preferably identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of guanine at PS1, guanine at PS2, cytosine at PS5, cytosine at PS6, thymine at PS7, adenine at PS8, adenine at PS9, adenine at PS10, thymine at PS11, adenine at PS12, thymine at PS13, adenine at PS14, guanine at PS16, thymine at PS17, thymine at PS18, cytosine at PS19, cytosine at PS20, thymine at PS21, guanine at PS22, adenine at PS23 and cytosine at PS24. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the CYP3A5 gene which is defined by any one of haplotypes 1- 11 and 13 - 26 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the CYP3A5 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant or fragment thereof, that is claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art. Any particular CYP3A5 variant or fragment thereof may also be prepared using synthetic or semi-synthetic methods known in the art.

CYP3A5 isogenes, or fragments thereof, may be isolated using any method that allows separation of the two "copies" of the CYP3A5 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No.

5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides CYP3A5 genome anthologies, which are collections of at least two CYP3A5 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same gender population. A CYP3A5 genome anthology may comprise individual CYP3A5 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the CYP3A5 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of such isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred CYP3A5 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded CYP3A5 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant CYP3A5 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as *E. coli*, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect,

eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the CYP3A5 gene will produce CYP3A5 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a CYP3A5 cDNA comprising a nucleotide sequence which is a polymorphic variant of the CYP3A5 reference coding sequence shown in Figure 2. Thus, the invention also provides CYP3A5 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2) (or its corresponding RNA sequence) for those regions of SEQ ID NO:2 that correspond to the examined portions of the CYP3A5 gene (as described in the Examples below), except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 88, adenine at a position corresponding to nucleotide 299 and guanine at a position corresponding to nucleotide 654, and may also comprise an additional polymorphism of adenine at a position corresponding to nucleotide 624. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a CYP3A5 isogene defined by any one of haplotypes 2, 5, 7-8, 18-19, and 21. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain one or more of the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified CYP3A5 mRNAs, cDNAs, or previously described fragments thereof. Polynucleotides comprising a variant CYP3A5 RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a CYP3A5 gene, mRNA or cDNA fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the CYP3A5 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the CYP3A5 gene or cDNA may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the CYP3A5 genomic, mRNA and

cDNA variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment of the invention may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular CYP3A5 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the CYP3A5 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular CYP3A5 isogene. Expression of a CYP3A5 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA or antisense RNA for the isogene or fragment thereof. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of CYP3A5 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of CYP3A5 mRNA transcribed from a particular isogene.

The untranslated mRNA, antisense RNA or antisense oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, such molecules may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of (a) the reference CYP3A5 amino acid sequence shown in Figure 3 or (b) a fragment of this reference sequence. The location of a variant amino acid in a CYP3A5 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A CYP3A5 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 for those regions of SEQ ID NO:3 that are encoded by examined portions of the CYP3A5 gene (as described in the Examples below), except for having one or more variant amino acids selected from the group consisting of tyrosine at a position corresponding to amino acid position 30 and tyrosine at a position corresponding to amino acid position 100. Thus, a CYP3A5 fragment of the invention, also referred to herein as a

CYP3A5 peptide variant, is any fragment of a CYP3A5 protein variant that contains one or more of the amino acid variations shown in Table 2. The invention specifically excludes amino acid sequences identical to those previously identified for CYP3A5, including SEQ ID NO:3, and previously described fragments thereof. CYP3A5 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a CYP3A5 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes, 2, 5, 7-8, 18-19, and 21, shown in Table 5.

Table 2. Novel Polymorphic Variants of CYP3A5

Polymorphic Variant Number	Amino Acid Position and Identities	
	30	100
1	H	Y
2	Y	S
3	Y	Y

A CYP3A5 peptide variant of the invention is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such CYP3A5 peptide variants may be useful as antigens to generate antibodies specific for one of the above CYP3A5 isoforms. In addition, the CYP3A5 peptide variants may be useful in drug screening assays.

A CYP3A5 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing an appropriate variant CYP3A5 genomic or cDNA sequence described above. Alternatively, the CYP3A5 protein variant may be isolated from a biological sample of an individual having a CYP3A5 isogene which encodes the variant protein. Where the sample contains two different CYP3A5 isoforms (i.e., the individual has different CYP3A5 isogenes), a particular CYP3A5 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular CYP3A5 isoform but does not bind to the other CYP3A5 isoform.

The expressed or isolated CYP3A5 protein or peptide may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the CYP3A5 protein or peptide as discussed further below. CYP3A5 variant proteins and peptides can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant CYP3A5 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric CYP3A5 protein. The non-CYP3A5 portion of the

chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the CYP3A5 and non-CYP3A5 portions so that the CYP3A5 protein may be cleaved and purified away from the non-CYP3A5 portion.

An additional embodiment of the invention relates to using a novel CYP3A5 protein isoform, or a fragment thereof, in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known CYP3A5 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The CYP3A5 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a CYP3A5 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the CYP3A5 protein(s) of interest and then washed. Bound CYP3A5 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel CYP3A5 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the CYP3A5 protein or to measure the enzymatic activity of CYP3A5 when using one or more candidate drugs as substrates.

In yet another embodiment, when a particular CYP3A5 haplotype or group of CYP3A5 haplotypes encodes a CYP3A5 protein variant with an amino acid sequence distinct from that of CYP3A5 protein isoforms encoded by other CYP3A5 haplotypes, then detection of that particular CYP3A5 haplotype or group of CYP3A5 haplotypes may be accomplished by detecting expression of the encoded CYP3A5 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel CYP3A5 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The CYP3A5 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the CYP3A5 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the CYP3A5 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the CYP3A5 protein variant from solution as well as react with CYP3A5 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect CYP3A5 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel CYP3A5 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the CYP3A5 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985; In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT patent applications, publication number WO 901443, and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of CYP3A5 may be

investigated by various means known in the art, such as by *in vitro* translation of mRNA transcripts of the CYP3A5 gene, cDNA or fragment thereof, or by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the CYP3A5 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA(s) into CYP3A5 protein(s) (including effects of polymorphisms on codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired CYP3A5 isogene, cDNA or coding sequence may be introduced into the cell in a vector such that the isogene, cDNA or coding sequence remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the CYP3A5 isogene, cDNA or coding sequence is introduced into a cell in such a way that it recombines with the endogenous CYP3A5 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired CYP3A5 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the CYP3A5 isogene, cDNA or coding sequence may be introduced include, but are not limited to, continuous culture cells, such as COS, CHO, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the CYP3A5 isogene, cDNA or coding sequence. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant CYP3A5 gene, cDNA or coding sequence are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene, cDNA or coding sequence is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes (or cDNA or coding sequence) of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the CYP3A5 isogene, cDNA or coding sequences may be introduced include, but are not limited to, mice, rats,

other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human CYP3A5 isogene, cDNA or coding sequence and producing the encoded human CYP3A5 protein can be used as biological models for studying diseases related to abnormal CYP3A5 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel CYP3A5 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel CYP3A5 isogenes (or cDNAs or coding sequences); an antisense oligonucleotide directed against one of the novel CYP3A5 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel CYP3A5 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel CYP3A5 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the

present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the CYP3A5 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The CYP3A5 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the CYP3A5 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the CYP3A5 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1).

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	3423-3448	complement of 3985-3960	563 nt
Fragment 2	3617-3639	complement of 4288-4266	672 nt
Fragment 3	3617-3639	complement of 4317-4294	701 nt
Fragment 4	7331-7353	complement of 7950-7928	620 nt
Fragment 5	9075-9098	complement of 9722-9703	648 nt
Fragment 6	11000-11022	complement of 11571-11550	572 nt
Fragment 7	16602-16626	complement of 17236-17214	635 nt
Fragment 8	16992-17013	complement of 17494-17474	503 nt
Fragment 9	18374-18395	complement of 18979-18957	606 nt
Fragment 10	19627-19650	complement of 20365-20340	739 nt
Fragment 11	20878-20900	complement of 21324-21302	447 nt
Fragment 12	23027-23049	complement of 23738-23715	712 nt
Fragment 13	30952-30975	complement of 31551-31528	600 nt
Fragment 14	33457-33479	complement of 34053-34032	597 nt
Fragment 15	35247-35271	complement of 35902-35878	656 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:

97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec.
72°C - 45 sec. } 10 cycles

97°C - 15 sec.
64°C - 45 sec.
72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in SEQ ID

NO:1 (Figure 1). Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	3456-3475	complement of 3960-3941
Fragment 2	3744-3764	complement of 4220-4201
Fragment 3	3744-3764	complement of 4286-4266
Fragment 4	7536-7557	complement of 7922-7902
Fragment 5	9202-9223	complement of 9594-9574
Fragment 6	11039-11058	complement of 11466-11447
Fragment 7	16655-16674	complement of 17183-17162
Fragment 8	17032-17052	complement of 17447-17427
Fragment 9	18403-18422	complement of 18950-18931
Fragment 10	19660-19679	complement of 20111-20090
Fragment 11	20904-20925	complement of 21264-21245
Fragment 12	23116-23137	complement of 23593-23572
Fragment 13	31065-31085	complement of 31451-31432
Fragment 14	33538-33558	complement of 33998-33977
Fragment 15	35308-35327	complement of 35849-35828

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the CYP3A5 reference genomic sequence (SEQ ID NO:1) are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the CYP3A5 Gene

Polymorphic Site Number	PolyId(a)	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	1225928	3633	A	G		
PS2	1225930	3747	C	G		
PS3(R)	1225932	3927	G	A		
PS4(R)	1225934	3939	C	T		
PS5	1225939	3998	A	C		
PS6	1225949	7657	T	C		
PS7	1225951	7717	C	T	88	H30Y
PS8	1225958	7830	G	A		
PS9	1225968	9523	T	A		
PS10	1225976	11189	C	A		
PS11	1225978	11214	C	T		
PS12	1225986	11310	C	A	299	S100Y
PS13	1226007	16830	C	T		
PS14	1226015	17383	G	A		
PS15(R)	1226017	18697	G	A	624	K208K
PS16	1226019	18727	A	G	654	P218P
PS17	1226021	18787	C	T		
PS18	1226023	19755	C	T		
PS19	1226027	19806	T	C		
PS20	1226029	20065	A	C		
PS21	1226033	21170	G	T		
PS22	1226035	31057	A	G		
PS23	1226037	33640	G	A		
PS24	1226041	35506	T	C		
PS25(R)	1226043	35618	T	C		

(a)PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.
(R)Reported previously

EXAMPLE 2

This example illustrates analysis of the CYP3A5 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in unrelated members of the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4 (Part 1). Genotypes and Haplotype Pairs Observed for CYP3A5 Gene

Genotype Number	HAP Pair		Polymorphic Sites PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10
1	12	12	A	C	G	C	A	T	C	G	T	C
2	15	15	A	C	G	C	A	T	C	G	T	C
3	11	11	A	C	G	C	A	T	C	G	T	C
4	12	4	A	C	G	C	A	T	C	G	T	C/A
5	12	22	A	C	G	C	A/C	T	C	G	T	C
6	11	20	A	C	G	C	A	T	C	G	T	C
7	12	17	A	C	G	C	A	T	C	G	T	C
8	12	19	A	C	G	C	A	T	C	G	T	C
9	12	16	A	C	G	C	A	T	C	G	T	C
10	12	5	A	C	G	C	A	T	C	G	T	C
11	12	6	A	C	G	C	A	T	C	G	T	C
12	11	15	A	C	G	C	A	T	C	G	T	C
13	12	8	A	C	G	C	A	T	C	G	T	C
14	12	23	A	C	G	C/T	A	T	C	G	T	C
15	14	13	A	C	G	C	A	T	C	G	T	C
16	12	20	A	C	G	C	A	T	C	G	T	C
17	11	7	A	C	G	C	A	T	C	G	T	C
18	12	21	A	C	G	C	A	T	C/T	G	T	C
19	11	25	A	C/G	G	C	A	T	C	G	T/A	C
20	11	2	A	C	G	C	A	T/C	C	G	T	C
21	11	3	A	C	G	C	A	T	C	G/A	T	C
22	12	24	A	C	G	C/T	A	T	C	G	T	C
23	11	18	A	C	G	C	A	T	C	G	T	C
24	12	1	A	C	G/A	C	A	T	C	G	T	C
25	12	9	A	C	G	C	A	T	C	G	T	C
26	12	14	A	C	G	C	A	T	C	G	T	C
27	12	26	A/G	C	G	C	A	T	C	G	T	C
28	15	8	A	C	G	C	A	T	C	G	T	C
29	12	15	A	C	G	C	A	T	C	G	T	C
30	12	10	A	C	G	C	A	T	C	G	T	C
31	12	11	A	C	G	C	A	T	C	G	T	C

Table 4 (Part 2). Genotypes and Haplotype Pairs Observed for CYP3A5 Gene

Genotype Number	HAP Pair		Polymorphic Sites PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20
1	12	12	C	C	C	G	G	A	C	C	T	A
2	15	15	C	C	C	G	G	A	C	C	T	A
3	11	11	C	C	C	G	G	A	C	C	T	A
4	12	4	C	C	C	G	G	A	C	C	T	A
5	12	22	C	C	C	G	G	A	C	C	T	A
6	11	20	C/T	C	C	G	G	A	C	C	T	A
7	12	17	C	C	C	G	G	A	C	C/T	T	A
8	12	19	C	C	C/T	G	G/A	A	C	C	T	A/C
9	12	16	C	C	C	G	G	A	C	C	T	A/C
10	12	5	C	C/A	C	G	G	A	C	C	T	A
11	12	6	C	C	C	G/A	G	A	C	C	T	A
12	11	15	C	C	C	G	G	A	C	C	T	A
13	12	8	C	C	C	G	G/A	A	C/T	C	T	A
14	12	23	C	C	C	G	G	A	C	C	T	A
15	14	13	C	C	C	G	G	A	C	C	T	A
16	12	20	C/T	C	C	G	G	A	C	C	T	A
17	11	7	C	C	C	G	G/A	A	C	C	T	A
18	12	21	C	C	C	G	G/A	A	C	C	T	A
19	11	25	C	C	C	G	G	A	C	C	T	A
20	11	2	C	C	C	G	G/A	A	C	C	T	A
21	11	3	C	C	C	G	G	A	C	C	T	A
22	12	24	C/T	C	C	G	G	A	C	C	T	A
23	11	18	C	C	C	G	G	A/G	C	C	T	A
24	12	1	C	C	C	G	G	A	C	C	T	A
25	12	9	C	C	C	G	G	A	C	C	T	A
26	12	14	C	C	C	G	G	A	C	C	T	A
27	12	26	C	C	C	G	G	A	C	C	T	A
28	15	8	C	C	C	G	G/A	A	C/T	C	T	A
29	12	15	C	C	C	G	G	A	C	C	T	A
30	12	10	C	C	C	G	G	A	C	C	T	A
31	12	11	C	C	C	G	G	A	C	C	T	A

Table 4 (Part 3). Genotypes and Haplotype Pairs Observed for CYP3A5 Gene

Genotype Number	HAP Pair		Polymorphic Sites PS21	PS22	PS23	PS24	PS25
1	12	12	G	A	G	T	T
2	15	15	T	A	G	T	C
3	11	11	G	A	G	T	C
4	12	4	G/T	A	G	T	T/C
5	12	22	G	A/G	G	T	T/C
6	11	20	G/T	A	G	T	C
7	12	17	G	A	G	T	T/C
8	12	19	G	A	G	T	T/C
9	12	16	G	A	G	T	T
10	12	5	G	A	G	T	T
11	12	6	G	A	G	T	T
12	11	15	G/T	A	G	T	C
13	12	8	G	A	G	T	T/C
14	12	23	G	A	G	T	T
15	14	13	G	G	G	T	T/C
16	12	20	G/T	A	G	T	T/C
17	11	7	G	A	G	T	C
18	12	21	G	A	G	T	T/C
19	11	25	G	A	G	T	C
20	11	2	G	A	G	T	C
21	11	3	G	A	G	T	C
22	12	24	G/T	A	G	T	T/C
23	11	18	G	A	G	T	C
24	12	1	G	A	G	T	T
25	12	9	G	A	G/A	T	T
26	12	14	G	A/G	G	T	T
27	12	26	G/T	A	G	T	T/C
28	15	8	T/G	A	G	T	C
29	12	15	G/T	A	G	T	T/C
30	12	10	G	A	G	T/C	T
31	12	11	G	A	G	T	T/C

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in PCT/US01/12831, filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In the present analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 26 human CYP3A5 haplotypes shown in Table 5 below.

A CYP3A5 isogene defined by a full-haplotype shown in Table 5 below comprises the regions

of the SEQ ID NOS indicated in Table 5, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 5.

Table 5 (Part 1). Haplotypes of the CYP3A5 gene.

Regions Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d) 1	2	3	4	5	6	7	8	9	10
3423-4317	1	3633/30	A	A	A	A	A	A	A	A	A	A
3423-4317	2	3747/150	C	C	C	C	C	C	C	C	C	C
3423-4317	3	3927/270	A	G	G	G	G	G	G	G	G	G
3423-4317	4	3939/390	C	C	C	C	C	C	C	C	C	C
3423-4317	5	3998/510	A	A	A	A	A	A	A	A	A	A
7331-7950	6	7657/630	T	C	T	T	T	T	T	T	T	T
7331-7950	7	7717/750	C	C	C	C	C	C	C	C	C	C
7331-7950	8	7830/870	G	G	A	G	G	G	G	G	G	G
9075-9722	9	9523/990	T	T	T	T	T	T	T	T	T	T
11000-11571	10	11189/1110	C	C	C	A	C	C	C	C	C	C
11000-11571	11	11214/1230	C	C	C	C	C	C	C	C	C	C
11000-11571	12	11310/1350	C	C	C	C	A	C	C	C	C	C
16602-17494	13	16830/1470	C	C	C	C	C	C	C	C	C	C
16602-17494	14	17383/1590	G	G	G	G	G	A	G	G	G	G
18374-18979	15	18697/1710	G	A	G	G	G	G	A	A	G	G
18374-18979	16	18727/1830	A	A	A	A	A	A	A	A	A	A
18374-18979	17	18787/1950	C	C	C	C	C	C	C	T	C	C
19627-20365	18	19755/2070	C	C	C	C	C	C	C	C	C	C
19627-20365	19	19806/2190	T	T	T	T	T	T	T	T	T	T
19627-20365	20	20065/2310	A	A	A	A	A	A	A	A	A	A
20878-21324	21	21170/2430	G	G	G	T	G	G	G	G	G	G
23027-23738	-	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057/2550	A	A	A	A	A	A	A	A	A	A
33457-34053	23	33640/2670	G	G	G	G	G	G	G	G	A	G
35247-35902	24	35506/2790	T	T	T	T	T	T	T	T	T	C
35247-35902	25	35618/2910	T	C	C	C	T	T	C	C	T	T

Table 5 (Part 2). Haplotypes of the CYP3A5 gene.

Regions Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d) 11	12	13	14	15	16	17	18	19	20
3423-4317	1	3633/30	A	A	A	A	A	A	A	A	A	A
3423-4317	2	3747/150	C	C	C	C	C	C	C	C	C	C
3423-4317	3	3927/270	G	G	G	G	G	G	G	G	G	G
3423-4317	4	3939/390	C	C	C	C	C	C	C	C	C	C
3423-4317	5	3998/510	A	A	A	A	A	A	A	A	A	A
7331-7950	6	7657/630	T	T	T	T	T	T	T	T	T	T
7331-7950	7	7717/750	C	C	C	C	C	C	C	C	C	C
7331-7950	8	7830/870	G	G	G	G	G	G	G	G	G	G
9075-9722	9	9523/990	T	T	T	T	T	T	T	T	T	T
11000-11571	10	11189/1110	C	C	C	C	C	C	C	C	C	C
11000-11571	11	11214/1230	C	C	C	C	C	C	C	C	C	T
11000-11571	12	11310/1350	C	C	C	C	C	C	C	C	C	C
16602-17494	13	16830/1470	C	C	C	C	C	C	C	C	T	C
16602-17494	14	17383/1590	G	G	G	G	G	G	G	G	G	G
18374-18979	15	18697/1710	G	G	G	G	G	G	G	G	A	G
18374-18979	16	18727/1830	A	A	A	A	A	A	A	G	A	A
18374-18979	17	18787/1950	C	C	C	C	C	C	C	C	C	C
19627-20365	18	19755/2070	C	C	C	C	C	C	T	C	C	C
19627-20365	19	19806/2190	T	T	T	T	T	T	T	T	T	T
19627-20365	20	20065/2310	A	A	A	A	A	C	A	A	C	A
20878-21324	21	21170/2430	G	G	G	G	T	G	G	G	G	T
23027-23738	-	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057/2550	A	A	G	G	A	A	A	A	A	A
33457-34053	23	33640/2670	G	G	G	G	G	G	G	G	G	G
35247-35902	24	35506/2790	T	T	T	T	T	T	T	T	T	T
35247-35902	25	35618/2910	C	T	C	T	C	T	C	C	C	C

Table 5 (Part 3). Haplotypes of the CYP3A5 gene.

Regions Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d) 21	22	23	24	25	26
3423-4317	1	3633/30	A	A	A	A	A	G
3423-4317	2	3747/150	C	C	C	C	G	C
3423-4317	3	3927/270	G	G	G	G	G	G
3423-4317	4	3939/390	C	C	T	T	C	C
3423-4317	5	3998/510	A	C	A	A	A	A
7331-7950	6	7657/630	T	T	T	T	T	T
7331-7950	7	7717/750	T	C	C	C	C	C
7331-7950	8	7830/870	G	G	G	G	G	G
9075-9722	9	9523/990	T	T	T	T	A	T
11000-11571	10	11189/1110	C	C	C	C	C	C
11000-11571	11	11214/1230	C	C	C	T	C	C
11000-11571	12	11310/1350	C	C	C	C	C	C
16602-17494	13	16830/1470	C	C	C	C	C	C
16602-17494	14	17383/1590	G	G	G	G	G	G
18374-18979	15	18697/1710	A	G	G	G	G	G
18374-18979	16	18727/1830	A	A	A	A	A	A
18374-18979	17	18787/1950	C	C	C	C	C	C
19627-20365	18	19755/2070	C	C	C	C	C	C
19627-20365	19	19806/2190	T	T	T	T	T	T
19627-20365	20	20065/2310	A	A	A	A	A	A
20878-21324	21	21170/2430	G	G	G	T	G	T
23027-23738	-	-	-	-	-	-	-	-
30952-31551	22	31057/2550	A	G	A	A	A	A
33457-34053	23	33640/2670	G	G	G	G	G	G
35247-35902	24	35506/2790	T	T	T	T	T	T
35247-35902	25	35618/2910	C	C	T	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within the indicated SEQ ID NO, with the 1st position number referring to SEQ ID NO:1 and the 2nd position number referring to SEQ ID NO:109, a modified version of SEQ ID NO:1 that comprises the context sequence of each polymorphic site, PS1-PS25, to facilitate electronic searching of the haplotypes;
(d) Alleles for CYP3A5 haplotypes are presented 5′ to 3′ in each column.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:109 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS25 in a uniform format to facilitate electronic searching of the CYP3A5 haplotypes. For each polymorphic site, SEQ ID NO:109 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 6 below shows the percent of chromosomes characterized by a given CYP3A5 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given CYP3A5 haplotype pair is shown in

Table 7. In Tables 6 and 7, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 6 and 7 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 6. Frequency of Observed CYP3A5 Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	1231283	0.61	2.38	0.0	0.0	0.0	0.0
2	1231274	0.61	0.0	2.5	0.0	0.0	0.0
3	1231279	0.61	0.0	2.5	0.0	0.0	0.0
4	1231280	0.61	0.0	0.0	0.0	2.78	0.0
5	1231287	0.61	2.38	0.0	0.0	0.0	0.0
6	1231286	0.61	0.0	0.0	0.0	2.78	0.0
7	1231266	1.83	0.0	7.5	0.0	0.0	0.0
8	1231267	1.22	0.0	0.0	0.0	5.56	0.0
9	1231285	0.61	0.0	0.0	2.5	0.0	0.0
10	1231284	0.61	0.0	2.5	0.0	0.0	0.0
11	1231263	9.76	0.0	37.5	0.0	2.78	0.0
12	1231262	59.76	73.81	27.5	67.5	66.67	83.33
13	1231282	0.61	2.38	0.0	0.0	0.0	0.0
14	1231265	6.1	14.29	2.5	0.0	5.56	16.67
15	1231264	7.32	0.0	2.5	22.5	5.56	0.0
16	1231271	0.61	2.38	0.0	0.0	0.0	0.0
17	1231281	0.61	0.0	0.0	2.5	0.0	0.0
18	1231269	1.22	0.0	5.0	0.0	0.0	0.0
19	1231277	0.61	0.0	0.0	2.5	0.0	0.0
20	1231268	1.22	2.38	2.5	0.0	0.0	0.0
21	1231275	0.61	0.0	2.5	0.0	0.0	0.0
22	1231273	0.61	0.0	0.0	0.0	2.78	0.0
23	1231270	1.22	0.0	0.0	0.0	5.56	0.0
24	1231278	0.61	0.0	2.5	0.0	0.0	0.0
25	1231276	0.61	0.0	2.5	0.0	0.0	0.0
26	1231272	0.61	0.0	0.0	2.5	0.0	0.0

Table 7. Frequency of Observed CYP3A5 Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
12	12	37.8	52.38	15.0	40.0	38.89	66.67
15	15	1.22	0.0	0.0	5.0	0.0	0.0
11	11	2.44	0.0	10.0	0.0	0.0	0.0
12	4	1.22	0.0	0.0	0.0	5.56	0.0
12	22	1.22	0.0	0.0	0.0	5.56	0.0
11	20	1.22	0.0	5.0	0.0	0.0	0.0
12	17	1.22	0.0	0.0	5.0	0.0	0.0
12	19	1.22	0.0	0.0	5.0	0.0	0.0
12	16	1.22	4.76	0.0	0.0	0.0	0.0
12	5	1.22	4.76	0.0	0.0	0.0	0.0
12	6	1.22	0.0	0.0	0.0	5.56	0.0
11	15	1.22	0.0	5.0	0.0	0.0	0.0
12	8	1.22	0.0	0.0	0.0	5.56	0.0
12	23	2.44	0.0	0.0	0.0	11.11	0.0
14	13	1.22	4.76	0.0	0.0	0.0	0.0
12	20	1.22	4.76	0.0	0.0	0.0	0.0
11	7	3.66	0.0	15.0	0.0	0.0	0.0
12	21	1.22	0.0	5.0	0.0	0.0	0.0
11	25	1.22	0.0	5.0	0.0	0.0	0.0
11	2	1.22	0.0	5.0	0.0	0.0	0.0
11	3	1.22	0.0	5.0	0.0	0.0	0.0
12	24	1.22	0.0	5.0	0.0	0.0	0.0
11	18	2.44	0.0	10.0	0.0	0.0	0.0
12	1	1.22	4.76	0.0	0.0	0.0	0.0
12	9	1.22	0.0	0.0	5.0	0.0	0.0
12	14	10.98	23.81	5.0	0.0	11.11	33.33
12	26	1.22	0.0	0.0	5.0	0.0	0.0
15	8	1.22	0.0	0.0	0.0	5.56	0.0
12	15	9.76	0.0	0.0	35.0	5.56	0.0
12	10	1.22	0.0	5.0	0.0	0.0	0.0
12	11	2.44	0.0	5.0	0.0	5.56	0.0

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the CYP3A5

gene are likely to be similar to the relative frequencies of these CYP3A5 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene of an individual, which comprises determining which of the CYP3A5 haplotypes shown in the table immediately below defines one copy of the individual's CYP3A5 gene, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS25 on at least one copy of the individual's CYP3A5 gene, and wherein each of the CYP3A5 haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1) 1	2	3	4	5	6	7	8	9	10
1	3633	A	A	A	A	A	A	A	A	A	A
2	3747	C	C	C	C	C	C	C	C	C	C
3	3927	A	G	G	G	G	G	G	G	G	G
4	3939	C	C	C	C	C	C	C	C	C	C
5	3998	A	A	A	A	A	A	A	A	A	A
6	7657	T	C	T	T	T	T	T	T	T	T
7	7717	C	C	C	C	C	C	C	C	C	C
8	7830	G	G	A	G	G	G	G	G	G	G
9	9523	T	T	T	T	T	T	T	T	T	T
10	11189	C	C	C	A	C	C	C	C	C	C
11	11214	C	C	C	C	C	C	C	C	C	C
12	11310	C	C	C	C	A	C	C	C	C	C
13	16830	C	C	C	C	C	C	C	C	C	C
14	17383	G	G	G	G	G	A	G	G	G	G
15	18697	G	A	G	G	G	G	A	A	G	G
16	18727	A	A	A	A	A	A	A	A	A	A
17	18787	C	C	C	C	C	C	C	T	C	C
18	19755	C	C	C	C	C	C	C	C	C	C
19	19806	T	T	T	T	T	T	T	T	T	T
20	20065	A	A	A	A	A	A	A	A	A	A
21	21170	G	G	G	T	G	G	G	G	G	G
22	31057	A	A	A	A	A	A	A	A	A	A
23	33640	G	G	G	G	G	G	G	G	A	G
24	35506	T	T	T	T	T	T	T	T	T	C
25	35618	T	C	C	C	T	T	C	C	T	T

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 2) 11	12	13	14	15	16	17	18	19	20
1	3633	A	A	A	A	A	A	A	A	A	A
2	3747	C	C	C	C	C	C	C	C	C	C
3	3927	G	G	G	G	G	G	G	G	G	G
4	3939	C	C	C	C	C	C	C	C	C	C
5	3998	A	A	A	A	A	A	A	A	A	A
6	7657	T	T	T	T	T	T	T	T	T	T
7	7717	C	C	C	C	C	C	C	C	C	C
8	7830	G	G	G	G	G	G	G	G	G	G
9	9523	T	T	T	T	T	T	T	T	T	T
10	11189	C	C	C	C	C	C	C	C	C	C
11	11214	C	C	C	C	C	C	C	C	C	T
12	11310	C	C	C	C	C	C	C	C	C	C
13	16830	C	C	C	C	C	C	C	C	T	C
14	17383	G	G	G	G	G	G	G	G	G	G
15	18697	G	G	G	G	G	G	G	G	A	G
16	18727	A	A	A	A	A	A	A	G	A	A
17	18787	C	C	C	C	C	C	C	C	C	C
18	19755	C	C	C	C	C	C	T	C	C	C
19	19806	T	T	T	T	T	T	T	T	T	T
20	20065	A	A	A	A	A	C	A	A	C	A
21	21170	G	G	G	G	T	G	G	G	G	T
22	31057	A	A	G	G	A	A	A	A	A	A
23	33640	G	G	G	G	G	G	G	G	G	G
24	35506	T	T	T	T	T	T	T	T	T	T
25	35618	C	T	C	T	C	T	C	C	C	C

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 3) 21	22	23	24	25	26
1	3633	A	A	A	A	A	G
2	3747	C	C	C	C	G	C
3	3927	G	G	G	G	G	G
4	3939	C	C	T	T	C	C
5	3998	A	C	A	A	A	A
6	7657	T	T	T	T	T	T
7	7717	T	C	C	C	C	C
8	7830	G	G	G	G	G	G
9	9523	T	T	T	T	A	T
10	11189	C	C	C	C	C	C
11	11214	C	C	C	T	C	C
12	11310	C	C	C	C	C	C
13	16830	C	C	C	C	C	C
14	17383	G	G	G	G	G	G
15	18697	A	G	G	G	G	G
16	18727	A	A	A	A	A	A
17	18787	C	C	C	C	C	C
18	19755	C	C	C	C	C	C
19	19806	T	T	T	T	T	T
20	20065	A	A	A	A	A	A
21	21170	G	G	G	T	G	T
22	31057	A	G	A	A	A	A
23	33640	G	G	G	G	G	G
24	35506	T	T	T	T	T	T
25	35618	C	C	T	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column.

2. A method for haplotyping the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene of an individual, which comprises determining which of the CYP3A5 haplotype pairs shown in the table immediately below defines both copies of the individual's CYP3A5 gene, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS25 on both copies of the individual's CYP3A5 gene, and wherein each of the CYP3A5 haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions and identities are set forth in the table immediately below:

PS	PS	Haplotype Pair(c) (Part 1)							
No.(a)	Position(b)	12/12	15/15	11/11	12/4	12/22	11/20	12/17	12/19
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/C	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/A	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C
21	21170	G/G	T/T	G/G	G/T	G/G	G/T	G/G	G/G
22	31057	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	T/T	C/C	C/C	T/C	T/C	C/C	T/C	T/C

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2) 12/16	12/5	12/6	11/15	12/8	12/23	14/13	12/20
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
12	11310	C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/C	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/T
22	31057	A/A	A/A	A/A	A/A	A/A	A/A	G/G	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	T/T	T/T	T/T	C/C	T/C	T/T	T/C	T/C

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3) 11/7	12/21	11/25	11/2	11/3	12/24	11/18	12/1
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T
7	7717	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
9	9523	T/T	T/T	T/A	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A
17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G
22	31057	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	C/C	T/C	C/C	C/C	C/C	T/C	C/C	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4) 12/9	12/14	12/26	15/8	12/15	12/10	12/11
1	3633	A/A	A/A	A/G	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/A	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/T	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/T	T/G	G/T	G/G	G/G
22	31057	A/A	A/G	A/A	A/A	A/A	A/A	A/A
23	33640	G/A	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/C	T/T
25	35618	T/T	T/T	T/C	C/C	T/C	T/T	T/C

(a) PS = polymorphic site;
(b)Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column.

3. A method for genotyping the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene of an individual, comprising determining for the two copies of the CYP3A5 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the one or more polymorphic sites (PS) have the position and alternative alleles shown in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the CYP3A5 gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for genotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the

hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

5. The method of claim 3, which comprises determining for the two copies of the CYP3A5 gene present in the individual the identity of the nucleotide pair at each of PS1-PS25.

6. A method for haplotyping the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene of an individual which comprises determining, for one copy of the CYP3A5 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

7. The method of claim 6, further comprising determining the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS3, PS4, PS15 and PS25, wherein the one or more polymorphic sites (PS) have the position and alternative alleles shown in SEQ ID NO:1.

8. The method of claim 6, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the CYP3A5 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid sample a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for haplotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

9. A method for predicting a haplotype pair for the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene of an individual comprising:

(a) identifying a CYP3A5 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1;

(b) comparing the genotype to the haplotype pair data set forth in the table immediately

below; and

(c) determining which haplotype pair is consistent with the genotype of the individual and with the haplotype pair data

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1) 12/12	15/15	11/11	12/4	12/22	11/20	12/17	12/19
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/C	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/A	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C
21	21170	G/G	T/T	G/G	G/T	G/G	G/T	G/G	G/G
22	31057	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	T/T	C/C	C/C	T/C	T/C	C/C	T/C	T/C

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2) 12/16	12/5	12/6	11/15	12/8	12/23	14/13	12/20
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
12	11310	C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/C	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/T
22	31057	A/A	A/A	A/A	A/A	A/A	A/A	G/G	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	T/T	T/T	T/T	C/C	T/C	T/T	T/C	T/C

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3) 11/7	12/21	11/25	11/2	11/3	12/24	11/18	12/1
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T
7	7717	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
9	9523	T/T	T/T	T/A	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A
17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G
22	31057	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	C/C	T/C	C/C	C/C	C/C	T/C	C/C	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4) 12/9	12/14	12/26	15/8	12/15	12/10	12/11
1	3633	A/A	A/A	A/G	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/A	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/T	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/T	T/G	G/T	G/G	G/G
22	31057	A/A	A/G	A/A	A/A	A/A	A/A	A/A
23	33640	G/A	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/C	T/T
25	35618	T/T	T/T	T/C	C/C	T/C	T/T	T/C

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column.

10. The method of claim 9, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS25, which have the position and alternative alleles shown in SEQ ID NO:1.

11. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-26 shown in the table presented immediately below, wherein each of the haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1) 1	2	3	4	5	6	7	8	9	10
1	3633	A	A	A	A	A	A	A	A	A	A
2	3747	C	C	C	C	C	C	C	C	C	C
3	3927	A	G	G	G	G	G	G	G	G	G
4	3939	C	C	C	C	C	C	C	C	C	C
5	3998	A	A	A	A	A	A	A	A	A	A
6	7657	T	C	T	T	T	T	T	T	T	T
7	7717	C	C	C	C	C	C	C	C	C	C
8	7830	G	G	A	G	G	G	G	G	G	G
9	9523	T	T	T	T	T	T	T	T	T	T
10	11189	C	C	C	A	C	C	C	C	C	C
11	11214	C	C	C	C	C	C	C	C	C	C
12	11310	C	C	C	C	A	C	C	C	C	C
13	16830	C	C	C	C	C	C	C	C	C	C
14	17383	G	G	G	G	G	A	G	G	G	G
15	18697	G	A	G	G	G	G	A	A	G	G
16	18727	A	A	A	A	A	A	A	A	A	A
17	18787	C	C	C	C	C	C	C	T	C	C
18	19755	C	C	C	C	C	C	C	C	C	C
19	19806	T	T	T	T	T	T	T	T	T	T
20	20065	A	A	A	A	A	A	A	A	A	A
21	21170	G	G	G	T	G	G	G	G	G	G
22	31057	A	A	A	A	A	A	A	A	A	A
23	33640	G	G	G	G	G	G	G	G	A	G
24	35506	T	T	T	T	T	T	T	T	T	C
25	35618	T	C	C	C	T	T	C	C	T	T

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 2) 11	12	13	14	15	16	17	18	19	20
1	3633	A	A	A	A	A	A	A	A	A	A
2	3747	C	C	C	C	C	C	C	C	C	C
3	3927	G	G	G	G	G	G	G	G	G	G
4	3939	C	C	C	C	C	C	C	C	C	C
5	3998	A	A	A	A	A	A	A	A	A	A
6	7657	T	T	T	T	T	T	T	T	T	T
7	7717	C	C	C	C	C	C	C	C	C	C
8	7830	G	G	G	G	G	G	G	G	G	G
9	9523	T	T	T	T	T	T	T	T	T	T
10	11189	C	C	C	C	C	C	C	C	C	C
11	11214	C	C	C	C	C	C	C	C	C	T
12	11310	C	C	C	C	C	C	C	C	C	C
13	16830	C	C	C	C	C	C	C	C	T	C
14	17383	G	G	G	G	G	G	G	G	G	G
15	18697	G	G	G	G	G	G	G	G	A	G
16	18727	A	A	A	A	A	A	A	G	A	A
17	18787	C	C	C	C	C	C	C	C	C	C
18	19755	C	C	C	C	C	C	T	C	C	C
19	19806	T	T	T	T	T	T	T	T	T	T
20	20065	A	A	A	A	A	C	A	A	C	A
21	21170	G	G	G	G	T	G	G	G	G	T
22	31057	A	A	G	G	A	A	A	A	A	A
23	33640	G	G	G	G	G	G	G	G	G	G
24	35506	T	T	T	T	T	T	T	T	T	T
25	35618	C	T	C	T	C	T	C	C	C	C

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 3) 21	22	23	24	25	26
1	3633	A	A	A	A	A	G
2	3747	C	C	C	C	G	C
3	3927	G	G	G	G	G	G
4	3939	C	C	T	T	C	C
5	3998	A	C	A	A	A	A
6	7657	T	T	T	T	T	T
7	7717	T	C	C	C	C	C
8	7830	G	G	G	G	G	G
9	9523	T	T	T	T	A	T
10	11189	C	C	C	C	C	C
11	11214	C	C	C	T	C	C
12	11310	C	C	C	C	C	C
13	16830	C	C	C	C	C	C
14	17383	G	G	G	G	G	G
15	18697	A	G	G	G	G	G
16	18727	A	A	A	A	A	A
17	18787	C	C	C	C	C	C
18	19755	C	C	C	C	C	C
19	19806	T	T	T	T	T	T
20	20065	A	A	A	A	A	A
21	21170	G	G	G	T	G	T
22	31057	A	G	A	A	A	A
23	33640	G	G	G	G	G	G
24	35506	T	T	T	T	T	T
25	35618	C	C	T	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the CYP3A5 haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions in SEQ ID NO:1 and identities are set forth in the table immediately below:

	PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1) 12/12	15/15	11/11	12/4	12/22	11/20	12/17	12/19
105	1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	5	3998	A/A	A/A	A/A	A/A	A/C	A/A	A/A	A/A
110	6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	10	11189	C/C	C/C	C/C	C/A	C/C	C/C	C/C	C/C
115	11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
	12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
	14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	15	18697	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
120	16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
	19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C
125	21	21170	G/G	T/T	G/G	G/T	G/G	G/T	G/G	G/G
	22	31057	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A
	23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	25	35618	T/T	C/C	C/C	T/C	T/C	C/C	T/C	T/C

	PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2) 12/16	12/5	12/6	11/15	12/8	12/23	14/13	12/20
	1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
135	3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
	5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
140	8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
	12	11310	C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C
145	13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	17383	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G
	15	18697	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
	16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	17	18787	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
150	18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	20	20065	A/C	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	21	21170	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/T
	22	31057	A/A	A/A	A/A	A/A	A/A	A/A	G/G	A/A
155	23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	25	35618	T/T	T/T	T/T	C/C	T/C	T/T	T/C	T/C

	PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3) 11/7	12/21	11/25	11/2	11/3	12/24	11/18	12/1
160	1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	2	3747	C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C
	3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
	4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
	5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
165	6	7657	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T
	7	7717	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
	8	7830	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
	9	9523	T/T	T/T	T/A	T/T	T/T	T/T	T/T	T/T
	10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
170	11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
	12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	15	18697	G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G
175	16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A
	17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
180	21	21170	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G
	22	31057	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	25	35618	C/C	T/C	C/C	C/C	C/C	T/C	C/C	T/T

185

	PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4) 12/9	12/14	12/26	15/8	12/15	12/10	12/11
	1	3633	A/A	A/A	A/G	A/A	A/A	A/A	A/A
	2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C
190	3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C
195	8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C
200	13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	15	18697	G/G	G/G	G/G	G/A	G/G	G/G	G/G
	16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	17	18787	C/C	C/C	C/C	C/T	C/C	C/C	C/C
205	18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	21	21170	G/G	G/G	G/T	T/G	G/T	G/G	G/G
	22	31057	A/A	A/G	A/A	A/A	A/A	A/A	A/A
210	23	33640	G/A	G/G	G/G	G/G	G/G	G/G	G/G
	24	35506	T/T	T/T	T/T	T/T	T/T	T/C	T/T
	25	35618	T/T	T/T	T/C	C/C	T/C	T/T	T/C

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

12. The method of claim 11, wherein the trait is a clinical response to a drug targeting or metabolized by CYP3A5 or to a drug for treating a condition or disease associated with CYP3A5 activity.

13. An isolated oligonucleotide designed for detecting a polymorphism in the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

14. The isolated oligonucleotide of claim 13, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the CYP3A5 gene at a region containing the polymorphic site.

15. The allele-specific oligonucleotide of claim 14, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-24, the complements of SEQ ID NOS:4-24, and SEQ ID NOS:25-66.

16. The isolated oligonucleotide of claim 13, which is a primer-extension oligonucleotide.

17. The primer-extension oligonucleotide of claim 16, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:67-108.

18. A kit for haplotyping or genotyping the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene of an individual, which comprises a set of oligonucleotides designed to haplotype or genotype each of polymorphic sites (PS) PS1, PS2, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

19. The kit of claim 18, which further comprises oligonucleotides designed to genotype or haplotype each of PS3, PS4, PS15 and PS25, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) isogene, wherein the CYP3A5 isogene is selected from the group consisting of isogenes 1- 11 and 13 - 26 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table

immediately below and wherein each of the isogenes 1- 11 and 13 - 26 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below; and

Region	PS	PS	Isogene Number(d) (Part 1)									
Examined(a)	No.(b)	Position(c)	1	2	3	4	5	6	7	8	9	10
3423-4317	1	3633	A	A	A	A	A	A	A	A	A	A
3423-4317	2	3747	C	C	C	C	C	C	C	C	C	C
3423-4317	3	3927	A	G	G	G	G	G	G	G	G	G
3423-4317	4	3939	C	C	C	C	C	C	C	C	C	C
3423-4317	5	3998	A	A	A	A	A	A	A	A	A	A
7331-7950	6	7657	T	C	T	T	T	T	T	T	T	T
7331-7950	7	7717	C	C	C	C	C	C	C	C	C	C
7331-7950	8	7830	G	G	A	G	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	T	T	T	T	T	T
11000-11571	10	11189	C	C	C	A	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	C	C	C	C	C	C	C
11000-11571	12	11310	C	C	C	C	A	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C	C	C	C	C
16602-17494	14	17383	G	G	G	G	G	A	G	G	G	G
18374-18979	15	18697	G	A	G	G	G	G	A	A	G	G
18374-18979	16	18727	A	A	A	A	A	A	A	A	A	A
18374-18979	17	18787	C	C	C	C	C	C	C	T	C	C
19627-20365	18	19755	C	C	C	C	C	C	C	C	C	C
19627-20365	19	19806	T	T	T	T	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	A	A	A	A	A
20878-21324	21	21170	G	G	G	T	G	G	G	G	G	G
23027-23738	-	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	A	A	A	A	A	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G	G	G	A	G
35247-35902	24	35506	T	T	T	T	T	T	T	T	T	C
35247-35902	25	35618	T	C	C	C	T	T	C	C	T	T

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2) 11	13	14	15	16	17	18	19	20
3423-4317	1	3633	A	A	A	A	A	A	A	A	A
3423-4317	2	3747	C	C	C	C	C	C	C	C	C
3423-4317	3	3927	G	G	G	G	G	G	G	G	G
3423-4317	4	3939	C	C	C	C	C	C	C	C	C
3423-4317	5	3998	A	A	A	A	A	A	A	A	A
7331-7950	6	7657	T	T	T	T	T	T	T	T	T
7331-7950	7	7717	C	C	C	C	C	C	C	C	C
7331-7950	8	7830	G	G	G	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	T	T	T	T	T
11000-11571	10	11189	C	C	C	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	C	C	C	C	C	T
11000-11571	12	11310	C	C	C	C	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C	C	T	C
16602-17494	14	17383	G	G	G	G	G	G	G	G	G
18374-18979	15	18697	G	G	G	G	G	G	G	A	G
18374-18979	16	18727	A	A	A	A	A	A	G	A	A
18374-18979	17	18787	C	C	C	C	C	C	C	C	C
19627-20365	18	19755	C	C	C	C	C	T	C	C	C
19627-20365	19	19806	T	T	T	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	C	A	A	C	A
20878-21324	21	21170	G	G	G	T	G	G	G	G	T
23027-23738	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	G	G	A	A	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G	G	G	G
35247-35902	24	35506	T	T	T	T	T	T	T	T	T
35247-35902	25	35618	C	C	T	C	T	C	C	C	C

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3) 21	22	23	24	25	26
3423-4317	1	3633	A	A	A	A	A	G
3423-4317	2	3747	C	C	C	C	G	C
3423-4317	3	3927	G	G	G	G	G	G
3423-4317	4	3939	C	C	T	T	C	C
3423-4317	5	3998	A	C	A	A	A	A
7331-7950	6	7657	T	T	T	T	T	T
7331-7950	7	7717	T	C	C	C	C	C
7331-7950	8	7830	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	A	T
11000-11571	10	11189	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	T	C	C
11000-11571	12	11310	C	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C
16602-17494	14	17383	G	G	G	G	G	G
18374-18979	15	18697	A	G	G	G	G	G
18374-18979	16	18727	A	A	A	A	A	A
18374-18979	17	18787	C	C	C	C	C	C
19627-20365	18	19755	C	C	C	C	C	C
19627-20365	19	19806	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	A
20878-21324	21	21170	G	G	G	T	G	T
23027-23738	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	G	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G
35247-35902	24	35506	T	T	T	T	T	T
35247-35902	25	35618	C	C	T	C	C	C

(a) Alleles for isogenes are presented 5′ to 3′ in each column;
(b) PS = polymorphic site;
(c) Position of PS in SEQ ID NO:1;
(d) Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

22. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 21, wherein the organism expresses a CYP3A5 protein that is encoded by the first nucleotide sequence.

23. The recombinant nonhuman organism of claim 22, which is a transgenic animal.

24. An isolated fragment of a cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) isogene, wherein the fragment comprises at least 10 nucleotides in one of the regions of SEQ ID NO:1 shown in the table immediately below and wherein the fragment comprises one or more polymorphisms selected from the group consisting of guanine at PS1, guanine at PS2, cytosine at PS5, cytosine at PS6, thymine at PS7, adenine at PS8, adenine at PS9, adenine at PS10,

thymine at PS11, adenine at PS12, thymine at PS13, adenine at PS14, guanine at PS16, thymine at PS17, thymine at PS18, cytosine at PS19, cytosine at PS20, thymine at PS21, guanine at PS22, adenine at PS23 and cytosine at PS24, wherein the selected polymorphism has the position set forth in the table immediately below:

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 1) 1	2	3	4	5	6	7	8	9	10
3423-4317	1	3633	A	A	A	A	A	A	A	A	A	A
3423-4317	2	3747	C	C	C	C	C	C	C	C	C	C
3423-4317	3	3927	A	G	G	G	G	G	G	G	G	G
3423-4317	4	3939	C	C	C	C	C	C	C	C	C	C
3423-4317	5	3998	A	A	A	A	A	A	A	A	A	A
7331-7950	6	7657	T	C	T	T	T	T	T	T	T	T
7331-7950	7	7717	C	C	C	C	C	C	C	C	C	C
7331-7950	8	7830	G	G	A	G	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	T	T	T	T	T	T
11000-11571	10	11189	C	C	C	A	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	C	C	C	C	C	C	C
11000-11571	12	11310	C	C	C	C	A	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C	C	C	C	C
16602-17494	14	17383	G	G	G	G	G	A	G	G	G	G
18374-18979	15	18697	G	A	G	G	G	G	A	A	G	G
18374-18979	16	18727	A	A	A	A	A	A	A	A	A	A
18374-18979	17	18787	C	C	C	C	C	C	C	T	C	C
19627-20365	18	19755	C	C	C	C	C	C	C	C	C	C
19627-20365	19	19806	T	T	T	T	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	A	A	A	A	A
20878-21324	21	21170	G	G	G	T	G	G	G	G	G	G
23027-23738	-	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	A	A	A	A	A	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G	G	G	A	G
35247-35902	24	35506	T	T	T	T	T	T	T	T	T	C
35247-35902	25	35618	T	C	C	C	T	T	C	C	T	T

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2) 11	13	14	15	16	17	18	19	20
3423-4317	1	3633	A	A	A	A	A	A	A	A	A
3423-4317	2	3747	C	C	C	C	C	C	C	C	C
3423-4317	3	3927	G	G	G	G	G	G	G	G	G
3423-4317	4	3939	C	C	C	C	C	C	C	C	C
3423-4317	5	3998	A	A	A	A	A	A	A	A	A
7331-7950	6	7657	T	T	T	T	T	T	T	T	T
7331-7950	7	7717	C	C	C	C	C	C	C	C	C
7331-7950	8	7830	G	G	G	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	T	T	T	T	T
11000-11571	10	11189	C	C	C	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	C	C	C	C	C	T
11000-11571	12	11310	C	C	C	C	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C	C	T	C
16602-17494	14	17383	G	G	G	G	G	G	G	G	G
18374-18979	15	18697	G	G	G	G	G	G	G	A	G
18374-18979	16	18727	A	A	A	A	A	A	G	A	A
18374-18979	17	18787	C	C	C	C	C	C	C	C	C
19627-20365	18	19755	C	C	C	C	C	T	C	C	C
19627-20365	19	19806	T	T	T	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	C	A	A	C	A
20878-21324	21	21170	G	G	G	T	G	G	G	G	T
23027-23738	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	G	G	A	A	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G	G	G	G
35247-35902	24	35506	T	T	T	T	T	T	T	T	T
35247-35902	25	35618	C	C	T	C	T	C	C	C	C

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3) 21	22	23	24	25	26
3423-4317	1	3633	A	A	A	A	A	G
3423-4317	2	3747	C	C	C	C	G	C
3423-4317	3	3927	G	G	G	G	G	G
3423-4317	4	3939	C	C	T	T	C	C
3423-4317	5	3998	A	C	A	A	A	A
7331-7950	6	7657	T	T	T	T	T	T
7331-7950	7	7717	T	C	C	C	C	C
7331-7950	8	7830	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	A	T
11000-11571	10	11189	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	T	C	C
11000-11571	12	11310	C	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C
16602-17494	14	17383	G	G	G	G	G	G
18374-18979	15	18697	A	G	G	G	G	G
18374-18979	16	18727	A	A	A	A	A	A
18374-18979	17	18787	C	C	C	C	C	C
19627-20365	18	19755	C	C	C	C	C	C
19627-20365	19	19806	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	A
20878-21324	21	21170	G	G	G	T	G	T
23027-23738	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	G	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G
35247-35902	24	35506	T	T	T	T	T	T
35247-35902	25	35618	C	C	T	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within SEQ ID NO:1;
(d) Alleles for CYP3A5 isogenes are presented 5′ to 3′ in each column.

25. An isolated polynucleotide comprising a coding sequence of a CYP3A5 isogene, wherein the coding sequence comprises SEQ ID NO:2, except at each of the polymorphic sites which have the positions in SEQ ID NO:2 and polymorphisms set forth in the table immediately below:

PS No.(a)	PS Position(b)	Isogene Coding Sequence Number(c) 2c	5c	7c	8c	18c	19c	21c
7	88	C	C	C	C	C	C	T
12	299	C	A	C	C	C	C	C
15	624	A	G	A	A	G	A	A
16	654	A	A	A	A	G	A	A

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:2;
(c) Alleles for the isogene coding sequence are presented 5′ to 3′ in each column; the numerical portion of the isogene coding sequence number represents the number of the parent full CYP3A5 isogene.

26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide

of claim 25, wherein the organism expresses a cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) protein that is encoded by the polymorphic variant sequence.

27. The recombinant nonhuman organism of claim 26, which is a transgenic animal.

28. An isolated fragment of a CYP3A5 coding sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 88, adenine at a position corresponding to nucleotide 299 and guanine at a position corresponding to nucleotide 654 in SEQ ID NO:2.

29 An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) protein, wherein the reference sequence comprises SEQ ID NO:3, except the polymorphic variant comprises one or more variant amino acids selected from the group consisting of tyrosine at a position corresponding to amino acid position 30 and tyrosine at a position corresponding to amino acid position 100.

30. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 29.

31. A method for screening for drugs, or other chemical compounds, that bind to or are enzymatic substrates for the isolated polypeptide of claim 29 which comprises contacting the CYP3A5 polymorphic variant with a candidate agent and assaying for binding activity.

32. An isolated fragment of a CYP3A5 protein, wherein the fragment comprises one or more variant amino acids selected from the group consisting of tyrosine at a position corresponding to amino acid position 30 and tyrosine at a position corresponding to amino acid position 100 in SEQ ID NO:3.

33. A computer system for storing and analyzing polymorphism data for the cytochrome P450, subfamily IIIA, polypeptide 5 gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

5 (d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises any one or more of the haplotypes set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1) 1	2	3	4	5	6	7	8	9	10
1	3633	A	A	A	A	A	A	A	A	A	A
2	3747	C	C	C	C	C	C	C	C	C	C
3	3927	A	G	G	G	G	G	G	G	G	G
4	3939	C	C	C	C	C	C	C	C	C	C
5	3998	A	A	A	A	A	A	A	A	A	A
6	7657	T	C	T	T	T	T	T	T	T	T
7	7717	C	C	C	C	C	C	C	C	C	C
8	7830	G	G	A	G	G	G	G	G	G	G
9	9523	T	T	T	T	T	T	T	T	T	T
10	11189	C	C	C	A	C	C	C	C	C	C
11	11214	C	C	C	C	C	C	C	C	C	C
12	11310	C	C	C	C	A	C	C	C	C	C
13	16830	C	C	C	C	C	C	C	C	C	C
14	17383	G	G	G	G	G	A	G	G	G	G
15	18697	G	A	G	G	G	G	A	A	G	G
16	18727	A	A	A	A	A	A	A	A	A	A
17	18787	C	C	C	C	C	C	C	T	C	C
18	19755	C	C	C	C	C	C	C	C	C	C
19	19806	T	T	T	T	T	T	T	T	T	T
20	20065	A	A	A	A	A	A	A	A	A	A
21	21170	G	G	G	T	G	G	G	G	G	G
22	31057	A	A	A	A	A	A	A	A	A	A
23	33640	G	G	G	G	G	G	G	G	A	G
24	35506	T	T	T	T	T	T	T	T	T	C
25	35618	T	C	C	C	T	T	C	C	T	T

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 2) 11	12	13	14	15	16	17	18	19	20
1	3633	A	A	A	A	A	A	A	A	A	A
2	3747	C	C	C	C	C	C	C	C	C	C
3	3927	G	G	G	G	G	G	G	G	G	G
4	3939	C	C	C	C	C	C	C	C	C	C
5	3998	A	A	A	A	A	A	A	A	A	A
6	7657	T	T	T	T	T	T	T	T	T	T
7	7717	C	C	C	C	C	C	C	C	C	C
8	7830	G	G	G	G	G	G	G	G	G	G
9	9523	T	T	T	T	T	T	T	T	T	T
10	11189	C	C	C	C	C	C	C	C	C	C
11	11214	C	C	C	C	C	C	C	C	C	T
12	11310	C	C	C	C	C	C	C	C	C	C
13	16830	C	C	C	C	C	C	C	C	T	C
14	17383	G	G	G	G	G	G	G	G	G	G
15	18697	G	G	G	G	G	G	G	G	A	G
16	18727	A	A	A	A	A	A	A	G	A	A
17	18787	C	C	C	C	C	C	C	C	C	C
18	19755	C	C	C	C	C	C	T	C	C	C
19	19806	T	T	T	T	T	T	T	T	T	T
20	20065	A	A	A	A	A	C	A	A	C	A
21	21170	G	G	G	G	T	G	G	G	G	T
22	31057	A	A	G	G	A	A	A	A	A	A
23	33640	G	G	G	G	G	G	G	G	G	G
24	35506	T	T	T	T	T	T	T	T	T	T
25	35618	C	T	C	T	C	T	C	C	C	C

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 3) 21	22	23	24	25	26
1	3633	A	A	A	A	A	G
2	3747	C	C	C	C	G	C
3	3927	G	G	G	G	G	G
4	3939	C	C	T	T	C	C
5	3998	A	C	A	A	A	A
6	7657	T	T	T	T	T	T
7	7717	T	C	C	C	C	C
8	7830	G	G	G	G	G	G
9	9523	T	T	T	T	A	T
10	11189	C	C	C	C	C	C
11	11214	C	C	C	T	C	C
12	11310	C	C	C	C	C	C
13	16830	C	C	C	C	C	C
14	17383	G	G	G	G	G	G
15	18697	A	G	G	G	G	G
16	18727	A	A	A	A	A	A
17	18787	C	C	C	C	C	C
18	19755	C	C	C	C	C	C
19	19806	T	T	T	T	T	T
20	20065	A	A	A	A	A	A
21	21170	G	G	G	T	G	T
22	31057	A	G	A	A	A	A
23	33640	G	G	G	G	G	G
24	35506	T	T	T	T	T	T
25	35618	C	C	T	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

the haplotype pairs set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1) 12/12	15/15	11/11	12/4	12/22	11/20	12/17	12/19
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/C	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/A	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C
21	21170	G/G	T/T	G/G	G/T	G/G	G/T	G/G	G/G
22	31057	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	T/T	C/C	C/C	T/C	T/C	C/C	T/C	T/C

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2) 12/16	12/5	12/6	11/15	12/8	12/23	14/13	12/20
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
12	11310	C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/C	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/T
22	31057	A/A	A/A	A/A	A/A	A/A	A/A	G/G	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	T/T	T/T	T/T	C/C	T/C	T/T	T/C	T/C

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3) 11/7	12/21	11/25	11/2	11/3	12/24	11/18	12/1
1	3633	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
4	3939	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T
7	7717	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G
9	9523	T/T	T/T	T/A	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A
17	18787	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G
22	31057	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
23	33640	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
25	35618	C/C	T/C	C/C	C/C	C/C	T/C	C/C	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4) 12/9	12/14	12/26	15/8	12/15	12/10	12/11
1	3633	A/A	A/A	A/G	A/A	A/A	A/A	A/A
2	3747	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	3927	G/G	G/G	G/G	G/G	G/G	G/G	G/G
4	3939	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	3998	A/A	A/A	A/A	A/A	A/A	A/A	A/A
6	7657	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	7717	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	7830	G/G	G/G	G/G	G/G	G/G	G/G	G/G
9	9523	T/T	T/T	T/T	T/T	T/T	T/T	T/T
10	11189	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	11214	C/C	C/C	C/C	C/C	C/C	C/C	C/C
12	11310	C/C	C/C	C/C	C/C	C/C	C/C	C/C
13	16830	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	17383	G/G	G/G	G/G	G/G	G/G	G/G	G/G
15	18697	G/G	G/G	G/G	G/A	G/G	G/G	G/G
16	18727	A/A	A/A	A/A	A/A	A/A	A/A	A/A
17	18787	C/C	C/C	C/C	C/T	C/C	C/C	C/C
18	19755	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	19806	T/T	T/T	T/T	T/T	T/T	T/T	T/T
20	20065	A/A	A/A	A/A	A/A	A/A	A/A	A/A
21	21170	G/G	G/G	G/T	T/G	G/T	G/G	G/G
22	31057	A/A	A/G	A/A	A/A	A/A	A/A	A/A
23	33640	G/A	G/G	G/G	G/G	G/G	G/G	G/G
24	35506	T/T	T/T	T/T	T/T	T/T	T/C	T/T
25	35618	T/T	T/T	T/C	C/C	T/C	T/T	T/C

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

and the frequency data in Tables 6 and 7.

34. A genome anthology for the cytochrome P450, subfamily IIIA, polypeptide 5 (CYP3A5) gene which comprises two or more CYP3A5 isogenes selected from the group consisting of isogenes 1-26 shown in the table immediately below, and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-26 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 1) 1	2	3	4	5	6	7	8	9	10
3423-4317	1	3633	A	A	A	A	A	A	A	A	A	A
3423-4317	2	3747	C	C	C	C	C	C	C	C	C	C
3423-4317	3	3927	A	G	G	G	G	G	G	G	G	G
3423-4317	4	3939	C	C	C	C	C	C	C	C	C	C
3423-4317	5	3998	A	A	A	A	A	A	A	A	A	A
7331-7950	6	7657	T	C	T	T	T	T	T	T	T	T
7331-7950	7	7717	C	C	C	C	C	C	C	C	C	C
7331-7950	8	7830	G	G	A	G	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	T	T	T	T	T	T
11000-11571	10	11189	C	C	C	A	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	C	C	C	C	C	C	C
11000-11571	12	11310	C	C	C	C	A	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C	C	C	C	C
16602-17494	14	17383	G	G	G	G	G	A	G	G	G	G
18374-18979	15	18697	G	A	G	G	G	G	A	A	G	G
18374-18979	16	18727	A	A	A	A	A	A	A	A	A	A
18374-18979	17	18787	C	C	C	C	C	C	C	T	C	C
19627-20365	18	19755	C	C	C	C	C	C	C	C	C	C
19627-20365	19	19806	T	T	T	T	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	A	A	A	A	A
20878-21324	21	21170	G	G	G	T	G	G	G	G	G	G
23027-23738	-	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	A	A	A	A	A	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G	G	G	A	G
35247-35902	24	35506	T	T	T	T	T	T	T	T	T	C
35247-35902	25	35618	T	C	C	C	T	T	C	C	T	T

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2) 11	12	13	14	15	16	17	18	19	20
3423-4317	1	3633	A	A	A	A	A	A	A	A	A	A
3423-4317	2	3747	C	C	C	C	C	C	C	C	C	C
3423-4317	3	3927	G	G	G	G	G	G	G	G	G	G
3423-4317	4	3939	C	C	C	C	C	C	C	C	C	C
3423-4317	5	3998	A	A	A	A	A	A	A	A	A	A
7331-7950	6	7657	T	T	T	T	T	T	T	T	T	T
7331-7950	7	7717	C	C	C	C	C	C	C	C	C	C
7331-7950	8	7830	G	G	G	G	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	T	T	T	T	T	T
11000-11571	10	11189	C	C	C	C	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	C	C	C	C	C	C	T
11000-11571	12	11310	C	C	C	C	C	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C	C	C	T	C
16602-17494	14	17383	G	G	G	G	G	G	G	G	G	G
18374-18979	15	18697	G	G	G	G	G	G	G	G	A	G
18374-18979	16	18727	A	A	A	A	A	A	A	G	A	A
18374-18979	17	18787	C	C	C	C	C	C	C	C	C	C
19627-20365	18	19755	C	C	C	C	C	C	T	C	C	C
19627-20365	19	19806	T	T	T	T	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	C	A	A	C	A
20878-21324	21	21170	G	G	G	G	T	G	G	G	G	T
23027-23738	-	-	-	-	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	A	G	G	A	A	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G	G	G	G	G
35247-35902	24	35506	T	T	T	T	T	T	T	T	T	T
35247-35902	25	35618	C	T	C	T	C	T	C	C	C	C

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3) 21	22	23	24	25	26
3423-4317	1	3633	A	A	A	A	A	G
3423-4317	2	3747	C	C	C	C	G	C
3423-4317	3	3927	G	G	G	G	G	G
3423-4317	4	3939	C	C	T	T	C	C
3423-4317	5	3998	A	C	A	A	A	A
7331-7950	6	7657	T	T	T	T	T	T
7331-7950	7	7717	T	C	C	C	C	C
7331-7950	8	7830	G	G	G	G	G	G
9075-9722	9	9523	T	T	T	T	A	T
11000-11571	10	11189	C	C	C	C	C	C
11000-11571	11	11214	C	C	C	T	C	C
11000-11571	12	11310	C	C	C	C	C	C
16602-17494	13	16830	C	C	C	C	C	C
16602-17494	14	17383	G	G	G	G	G	G
18374-18979	15	18697	A	G	G	G	G	G
18374-18979	16	18727	A	A	A	A	A	A
18374-18979	17	18787	C	C	C	C	C	C
19627-20365	18	19755	C	C	C	C	C	C
19627-20365	19	19806	T	T	T	T	T	T
19627-20365	20	20065	A	A	A	A	A	A
20878-21324	21	21170	G	G	G	T	G	T
23027-23738	-	-	-	-	-	-	-	-
30952-31551	22	31057	A	G	A	A	A	A
33457-34053	23	33640	G	G	G	G	G	G
35247-35902	24	35506	T	T	T	T	T	T
35247-35902	25	35618	C	C	T	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within SEQ ID NO:1;
(d) Alleles for CYP3A5 isogenes are presented 5′ to 3′ in each column.

POLYMORPHISMS IN THE CYP3A5 GENE

```
TTACTTTCCC TTCCTGAGTA ACTTATCCTA AAGTCATTAG GTGGGTGGCA
GCCAGATGGT GGCCACACAT TAAGGTAGAA AAGAGAGTGT CATGATGGTT      100
CCAAGTCAGA GACCTAGTAG GGTGAGGATC AAGTAGGTGT TCACGTGGAG
AAACAGCCCG GCCTGTGTGT GGGAGTCCAA GCAAGCAGAG AAAATGTCGA      200
CACAGAGGGG TGGCCTGAAA AAGCAGCCAG AGCCTAAACA GGGCATGGAG
AACATATTTA GGGCATGAGG TGAGGAGGGC ATCCATGAGT GGGAAGGGAT      300
GGGTGAGGTT TCACTACATA AAGGGGATTG ATGAAATAAG TAAATAAAGT
ATACTGGAAG CCAGGTGTGT CACTTTTGCA GAAAAGAGTC ATGGATTCAG      400
AAAGGGAGAA AACTAGCAGG AATCCTATGA AATTAGATTA AAATGGATGT
ATCCATGTAT ATTCATACCC TTCTAGATAG ATAAATGGTT AGATAGGTGA      500
TAAAAAGATA ACAAGAGGAC AAGATAATTA GATAGACATA AATGTATGTA
TGTGTTTGTG TGTGTGTACA AAAAAACATA TACTCCCTAC TTCTCTCCAC      600
TGATAGGGCT AGGTAACAAT GGCATTTCAA TAGCAATGAG CACACTTAGT
GGCCAGATCT TGGCTTATTA ATACCATTTT CCACTGAAAG GAACCAGAGC      700
TTTTTAGAGA AATGGCTGAT TCCAGGGCCA GGATTAAGAA TGTTCAAGAT
AAGCCTAGGA TACATTTTGT GCCAGGAAGC AAGAAGATGT TCAAATGATT      800
TCCAAGTAAT GTTTGGAAAT GATATTTGAA AATGATTTCC AAATGATATT
TCCAAATGAT TTCCAAATGA TATATGGAAA CACTTAAAGA CTCCACTAAA      900
GAACTATTAG ATCTGATAAA CAAATTCAGT AATGTTGCTG GATACAAAAT
CAACATACAA AAACCAGTAG CATTTCTGCA TGCCAACAGT GAACAATCTG     1000
GCAAAAATAA AAAATGTAAT CCCATTTACA ATAACCCCAA ATAAAACTAA
ATACCTGGGA ATTAACTTAA GAGAAAGATG TCTACAATTA ATATTGTAAA     1100
ACACTGATGA AGGAAATTGA AGAAGACACA AAAAAGAAGG ATATTCCATG
TTTATATATT GTAAGCATTA ATATTGTTAA AAATGTCCAT ACTACCCAAA     1200
GCAATGCACA GATTCAATGC AGTCTCTCAA AATACCAATG GCATTCTTCA
AAGAAATAGA AAAAAAAAAA CCCTAAAATT TGTATGGAAC CACAAAAGAC     1300
CCAGAATAGC GAAAGCTACC TTCAGCAAAA AGAACAAAAC TGGAGGAATC
ATATTACCTG ACTTCAAATT ATACTACAGA GGTATAATAA CCAAAACAGT     1400
ATGGTACTTG TATAAAAACA GACACAGACC AATGAAATAG AATAGAGAAC
CCAGAAACAA TTCCACACAC CTACGGTGAA CTCATTTTCA ACAATGTTGT     1500
CAAGAACATA CACTGGGGGA AAAGACAGTC TCTTCTGGTG CTGGGAAAGC
TGGATTTTAA CATGCAGAAT AATGAAACTA GAACCCTGTA TCTCACCAGA     1600
CACAAAAATC AAATCAAGGT GGACGAAAGA CTGAAACCTG GCTGAGTGCC
GTGGCTCATG CCTGTAATCC CAGCATTTTG AGAGGCCGAG GCGGGTGTAT     1700
CACTTGAGGT CAGGAGTTCA AGACCAGCCT GGCCAACATG GTGAAACCAC
ATGTCTACCA AAAAATACAA GAGTTAGCTG GACATGCTGG TGCGTGCCTG     1800
TAGTCCCAGC TACACAGAAG GCTGAGGTGG CAGAATCACT TGGACCCAGG
AGGCGGAGGT GGCAGTGAGC TGAGATCATG ACAATGCACC CCAGCCTGGG     1900
CAAGAGAGTG AGACTCTGTC AGAAAAACAA AAAACAAAAA AACAAAAAAC
AAAACTGAAA TCTGAGACCT CAAACGATGA AACTGCTACA AGAAAACATT     2000
GTGGAAACTC TTCAGGATAT TGGTCTGGGC AAAACTTTCT GAAGAACTAC
CCCACAAGCA CAGGCAACCA AAGCAAAAAT GGACAAATGG ATCAGATCAA     2100
GTTAAAAAGC TTCTGTACCA CAAAGAAAGC AATCAACAAA GTGAAGACAC
AAACCACAGA ATGGGAGAAA ATATTTTCAA AGTCACACTC TGACAACAGA     2200
TTAATAGCCA GAATACATGA AGCGCTCAAA CAACTCTGTA AGGAAAAATC
TAATAATCCA ATCAAAAAAT GGGCAAAATT TGAATAGACA TTTTTCAAAA     2300
GAAGACATAC AAATGCCACA TAGGCATATG ATAAGGTGCT CAACATCACT
GGTCATTAGA GAAATGCAAA TCAAAACCAC AATGAGATAT CATCTTACCC     2400
CAGCTAAAAT GGTTTTTATC CAAAAGACAG GCAACAACAA ATGCCAGCGA
GAATGTGGAG AAAAGGGAAC CCTTGTACAC TGTTGGTGTA AATTAGTGCA     2500
ACCACTATAG AGAACAATTT GGAGGTTCCT CAAAACATTA AAATTAACAT
```

FIGURE 1A

```
TAAATAGAGC TACCACAATA TCCAGAAATC CCCATGCTGG GTATATACCT      2600
GGAAGAAAGG AAATCATATA TTGAAGAGAT AACATCACTC CAATATTCAC
AATAGCCACT ATTCACAAAT GCCAAGATTT GGAAGCAACC TAAGTGTCCA      2700
TCAACAGATG AATGGATAAA GAAAGTACTC CAATTATACA CAATGGAGCA
CAATTCAGCC ATGAAAAAAG CATGAGATCC TGTTATCTGT AATAATATGG      2800
ATGGAACTGG AGGTCATCAT GTTAAGTGAA ATAAGCCAGG CACAGAAACA
CAGATATTGC AAGTTCTCAC ATACTTGTGG GATCTACAAA TCAAAACAAC      2900
TGAGCTAATG TCTGGGCCTT AGTCAGTGTT GTACCCAAGT ACTGGGAGCA
CAGCTTTTAA AATACATCAT GAATGCTTTA ATACAGGAAT GAATAGATGA      3000
GAGGCACAAA CTGGTTGGGT GTTCTTCTGA TACACAGTAT CTTCCTTGAC
AGATTCAGTA CAACTCTCAA CAGGTAAGTC TCTTCATGTT ATGTTACCTT      3100
ATGAGGAATT AAGTGGCAGA ACATGATTTC TATTATTTTC CTTTGCAGAA
CAAGACCAAC TTTATTAGTT GGGACACAGT GTGGCTGCAT TTGAGTCCCA      3200
AGCAACCATT AGTCTATTGC TATCACCACA GAGTCAGAGG GGATGAGACG
CCCAGCAATC TCACCCAAGA CAACTCCACC AACATTCCTG GTTACCCACC      3300
ATGTGTACAG TACCCTGCTA GGAACCAGGG TCATGAAAGT AAATAATACC
AGACTGTGCC CTTGAGGAGC TCACCTCTGC TAAGGGAAAC AGGCATAGAA      3400
ACTTACAATG GTGGTAGAGA GAAAAGAGGA CAATAGGACT GTGTGAGGGG
GATAGGAGGC ACCCAGAGGA GGAAATGGTT ACATTTGTGT GAGGAGGTTG      3500
GTAAGGAAAA ATTTTAGCAG AAGGGGTCTG TCTGGCTGGG CTTGGAAGGA
TACGTAGGAG TCATCTAGAG GGCACAGGTA CACTCCAGGC AGAGGGAATT      3600
TCGTGGGTAA AGATGTGTAG GTGTGGCTTG TGAGGATGGA TTTCAATTAT
                                  G
TCTAGAATGA AGGCAGCCAT GGAGGGGCAG GTGAGAGGAG GGTTAATAGA      3700
TTTCATGCCA ATGGCTCCAC TTGAGTTTCT GATAAGAACC CAGAACCCTT
                                                    G
GGACTCCCCG ATAACACTGA TTAAGCTTTT CATGATTCCT CATAGAACAT      3800
GAACTCAAAA GAGGTCAGCA AAGGGGTGTG TGCGATTCTT TGCTATTGGC
TGCAGCTATA GCCCTGCCTC CTTCTCCAGC ACATAAATCT TTCAGCAGCT      3900
TGGCTGAAGA CTGCTGTGCA GGGCAGGGAA GCTCCAGGCA AACAGCCCAG
                          A               T
CAAACAGCAG CACTCAGCTA AAAGGAAGAC TCACAGAACA CAGTTGAAGA      4000
                                                    C
AGGAAAGTGG CGATGGACCT CATCCCAAAT TTGGCGGTGG AAACCTGGCT
     [exon 1: 4013..
TCTCCTGGCT GTCAGCCTGG TGCTCCTCTA TCTGTGAGTA ACTGTCCAAA      4100
              ..4083]
CTCCTCTCTT TGTTTCCTTG GACTTGGGGT GCTAATCGGG CCCCTTTTCC
CTTATCTGTT TTGAAGATCA AAAGAGATGT TCAAGGAGAA GTAGCTGAAG      4200
TGTTGGACGC TACAAACGCA TAGAAGTTAT TATTATCTTA TGCAGATCTA
TGAATGAATA AATAAGCATT TCTCCCATCC ACCTTCTAAT TTTGGTGACT      4300
AGGAGGGTTT AGGGACAGCA TTTGGTAGTG GGAATGATTT GATTAGCTTA
GATCTGACGA AGACTAATCA ATGAAAACAT GGCAGCGGCA GATTACAAAC      4400
TGCTGATCAT GATGGACAGT GTGATCCTCA TCCCCTTCCC AGGCTCTGGG
GATTCTGGGT ACAGGAAGGA GTGGCTTGCA TTTTTGTCTC ATTAATTCGC      4500
TTTCTGGGTT CTGTGTCTGC TGGAAGGGAT GTGTAGCTGT ATTGCCCCTG
TAGACCTGGT TCCTGCTCCC CCGCCTTCCA ACCCAGGATA TCATTTACAT      4600
AACGCACCAG GGGACACCAA GACTTCATGG GAAGCTGTCC CCTGGCTCTT
CCCTCTTTCC TGTGCCATGC CCCTGAAAAT CCCCTCCCTC CTATGAGTCA      4700
CTCCTCCACC CTGTCATACA CAGGATGGTT TATCTTGCAA TGATTAACCT
CTAGAGCAAA GGAGACCTGG AGGAAGTTTC GAGGATTTAT TCTTTGCTTT      4800
AATCTTTTTC CTCCCGTCTC TGGGAGGCTA GGATTAATAT AGAGCTTTGT
TTCTCACCTA ATGGGAATCT ACTAGCAGCC TGAAAAGGCA GGAGCCATGA      4900
```

FIGURE 1B

```
AAGCCAATTT GGATTTTACA TATTTTTCCC CTTTATGTTA CAGTACAGGA
GGGCAAACCC TCTCACTGGT GGGATTCCTG GCATCCTAGA GCAGGTGGAG      5000
AGAAGAGTTA CTTTCCACTG TGGGTAGTGG AGGCTCCACC TGTCCCATTA
ACTTCTACCT CAATTTGACT TTTATTAAGA GCAGGGAACC ACAATGACAT      5100
GAAAATAGAC ACTATAAACC TCATTTTAAT TCTTTCACAG AAAGCTTAGG
AATTCAGTGA GTTGTGGCAA CATGGTTTCC ATTGTCTAAC ATTTTTAAAT      5200
GAATTGATAT GGTTTAAATT CATTCATTTT TAAACCAGAA TTTTTTGGAG
ATAGACTATT TCCAGCATGT TCCTTCTGGA TGGTAAAACA GGGCTGTTAG      5300
TTCAGTATTT GTGACAATAA GTGTGTGTAA AATAATGTCA CCTTTCCTGA
ATGTCAGGAA TATGAGTCTA ATGCACAAAT GTATACCTCT AAGACAAGAC      5400
TGCACGTCTT TTCAAATATA CCTGTCCGGC CATTTATTTT AATAACTCCT
TTTCGAATAT ACCTGCTTAG CAGATTGTCT TAAACTCTCA GGACAGGGGA      5500
GTAAGCAAGA CTGTGAGCCA GTGACGATAG CAAAGGCTTC CAGGTAGGAT
CCATATGAAG TGAGAAAATA TTCCTCAGCT CTCAGGGTAG AACTCCAAAG      5600
AGATATTCAT GGGTCCTGGC CCCACCGTGG AGGTCACTCA AAGGGCAAAC
AGGTTGGCAT CTCATCTGCT TCAAGCCTGG ACACAGGGGC ACCATCTGTG      5700
TCACTCTGTG TGTGGTCTGC CATGTTGTGG GCCGGTCACT ACAGACTCGG
GCAGCCAGGC AGACAATGCC TTAGCCTTAG ACAATGCTGG TGCAGCCCAG      5800
GAGTCAGAAA ATGCAGTGTA GACCAGGCCC TCCTTAGGCC AACACAATTA
CATGCAATAG ATGACTGGCT TTTCTGTTAG TCTCTTCACT GGACCCAAAG      5900
GCTGCATTAC TCTACCAGAG GGGAGCTGGA AAGAAACTAA AGAGTTCGCC
CAGCACAGCA TCTGCCTTGA CATGGTACCA TGTGAATCTA GACACTCACC      6000
AAGATCTTTC CTTGGGGGCC AATGCTGCTG ACACATTAAC TCAATAGCTT
GTCCTCACCT GAGAGGTCAG GTAATGTGTT TAAAGTTCAG GAGCAGAGAT      6100
TAGTGTCATT GATTTGACAT GGCTGTGACA ACAAAGGAGG GAACTGAAGT
GGGAATACCC AAGGCCACCC TGGCTTTGGC AGGTGGTGCA CGCACTTCCA      6200
CTAACTGTTC TGGGGCAGGG AACCAAATGT ATGACTGGGC CTGCTCATGC
TGCCCCTGCT GAGTCCTCCA AACCCTGCCC TTCATGTAAT TTCTCAGTTT      6300
TATTTTATCA CATTTTATAA GTCACTGGAT GTTTACAAAA TGTTTGGAAC
CTATACTGCC TTGAAGGCTA ACCTCTAAAG AGGAGTAAAC AAGGTCTTAA      6400
TACAACTCTC CGGGACGTTT TATCATTACT TATCTTATAT GCCATACTGC
ACCATTTGCT ATCAACAGGA AAGTACCTGG ACTTTGGAAG GTCCCTCTGT      6500
GTCTTTTAGC TGAAAGTACA TATGAGGCAT GTGGATTCTT TTATGCACAT
CATCTTTTTC AGCCACATTT TTGTAGTTTG CCTCTCTGGA GCCAACTGTG      6600
TGGGGCTAGC AGCTTCACAG CTGAATCAGT GTCTGGCAAC CTCTTCCTTC
AGCCTCTCTT CTTCCTCCAG TTTTCCATCC CTCAGTCACA CCGGAGGGGG      6700
AAGGTCTGCA AGGATCCAGA ACCATCAGTT GGAGGAGTTT GCACATGACT
CATGAAAGAT GAGTTCCAGG CAGGCCTGCC ATAGTGAACA CCAGGCTTAA      6800
TGGGTTTTTC CTCAGAGATA CTTCACGTAC AGAGGCAGTG AACTGACTGC
TTTCTGGTTG ACCACCTTGA AAAAGATGAG TGTGCCTGGC ACTGTGCTTC      6900
TCAGGTGAGT ATGACCTGAG AAGTATTAGT TGCTGGTTCT TCTGCACACA
ATCATTCAAG GACATATGGA TCAACCATCC TCCTCAACAG CTCAAATCAA      7000
CCAGATCATC TGACCACAGA GACTGAGGTG TACCTGAAAG CTGCCCACAT
TTCTATAAGG CCAATAGAAG CCATGAACAC AGTTGTCAAT CTGTAGAAAT      7100
AAGGACTCCA TGACTCCTCC AAGGCCTCTC TGTGAATGAA CGTTTAAGAA
GGGCTAGATC CTAAAACAGG GTCAGAGCTT AGAGGGAAGA AAAAGCATAA      7200
ACATTTCTGA GCAAATTGTA AGGGCAGTGT CACCATAGGC TCCCAGTGAC
CCTCTGTGAT TGAGTGCATA CAGTGATGCA AAATCTCATC ATCAGTGCAA      7300
AAGACAAAAA AAATCTTACT CTTTCTACCT AGGATGAGAG TCCCCAAATC
AGCGAAGAGT CCACTTACTA AACAGACATA AGGAAATGAA GTGTCCTGGA      7400
AGAATTCCTG CCTGAACCTC TCAGGAGCAT TTGAGGACAT TTATCAAGTA
TTCACTCCAG GATTGGGACT ATGAAGACTT CAGCTGCTTT CAGCTAATCA      7500
TTGAGACTTT TCAGGGGTCT CAGAATAGTC AGGAAAGGAC CTGATGAGTG
```

FIGURE 1C

```
AATGCAATTA CTGATGTTGG AGTTGCTGTT ATTATTTATC GTGTACATAT      7600
TACCTCCCTC TCTTGACCAT TCCAGTTCCT GAGTAACTCA CCAGCCCTCT
GATCTATAAA GTCACAATCC CTGTGACCTG ATTTCTGTTT CACTTTGTAG      7700
      C
ATATGGGACC CGTACACATG GACTTTTTAA GAGACTGGGA ATTCCAGGGC
                T
     [exon 2: 7701..
CCACACCTCT GCCTTTGTTG GGAAATGTTT TGTCCTATCG TCAGGTGAGT      7800
               ..7794]
TGCTTGAGCT TCCTCTTTTG CTTCTTATGG TTGCAAACAT CAGCTTAGTT
                              A
CCATCAGTAA AAATGCCCCT CCTTGGGAGG GAGTTCTGAG GTTTCACATT      7900
TTCAGAAATG GTGGGACTGG GTGCAGTGGA TCATGCCTGT AATCTCAGCC
TCTGTGAGGC CAAGACTGGC AAATTGCTTG AGCCCAGGAG TTTGAGAACA      8000
GCCTGGGCAA CACAGTGAGA CACCTGTCTC TAGAAAGAAA AAATTACCTG
TGCATGATAT GGTAGCCCAT GCCTGTAGTC CCAGCTACTC TGAATGTTAA      8100
GGTGGGAGGA TTGTATGAAC CCAGGAAGTC AAGGCTGTAT TGAGCTGTGA
TCGCACCACT GCACTCCAGC TTGGTCAACA GAACAAGACA GAAAGGAAGA      8200
AAGAAAGAGA GAGAGAGAAA GAAAGAGAGA GGAAGGAGAG GGGAGGGGAG
GGGAGGGGAG GGGGGAGGAG AGGAGAGGAG AGAAAAGGAG AGGAGAGAGG      8300
AGAGGAGAGG AAAAGGTGTG TAGGCTCCAC CCAAAGCATG GCCAGGTTTA
CCCCTGGAGG GAAAGTCACA AGCTCATGTC CAGAAGGCCA GTAGCAGCAA      8400
GCTGCTCTCC AGCCCAGATT TCCTATCCTG TGTACCTGGA GCTTGTTTCT
CAGATTCTAA CTCTCACAAC TGAAGCCTCT GTTGTCTGAT TACTATCTGA      8500
GAATTCTACA CAATTTTACC CTCGATAAAA GCAGTAATTT CTTCTTCATC
TTTCCCAGAT CAACTCTTGT AGTAGATCAA CATTTCTGGG ACCTTCTTTT      8600
GCATGGTTAA AACATCACAG CTGAATCTTA GCAACAGGAA GGTTTGTTTT
TATGTTTCAG AAGTGAAAGC TCAGAGCACG CATTGTAATT TGCTGGGTGT      8700
GATGTGTAGA GGTGGCATTT CTCCATCTTT TCTGTGTTAA GCTAGAAAAC
TGGAAAGGAA GTCTACTTTC TCATTCACTC ACTCACTTTC TCACTCAACA      8800
ACATGCCTTA GACTTATCTA AATCTGCAAG ACTAAAAGAG GTTCCTGGTT
TCTTTAACTT TCTAATTCTG CTAGAGTTCT AGAGAGAGCA CATGAGATAA      8900
ATGAAAAGGA TACTGATGGA GGAGATTAAA AAATTGTGCA TTCCCTGCAG
ACACTCACTT TTCCTCACCT CAGTTTCACC CCTGCCCTTG CAGGTGATCA      9000
TTCACGGGGT TAGGAGACTT TAGAGAGAAT AAAAGAAAAA GCAAAAATAC
ATCAGAAAGA CAAGGAATTA CTTACTGGTC ATAGACAAGG GTGAGTCCTT      9100
CAGTACTTAG AGAAAATTCA AGAGTGACTT TAAATTCCCC ACTTCAAATA
TATTCTCTGT TTTCTTGTCT TTCCCTTAAG ACATCTCTGA ATAGCTTCCT      9200
TCAACTGCCA GTGAAAGATA GCAGGCCTGA TTTCATTGGA CGCAACTGTT
TTCAGCCCCA ATTAGAGGTA GGGTTTATTC TATTTAAAAT AATAATCAAC      9300
TTGTATTTTG TTTCCTCTCC CAGGGTCTCT GGAAATTTGA CACAGAGTGC
     [exon 3: 9324..
TATAAAAAGT ATGGAAAAAT GTGGGGGTGA GTATTCTGAA AACCTCCATT      9400
               ..9376]
GGATAGACCT GCTACTGTGA GGAGGTTACC CCACTGCAGG ATAGTCTCTG
CCCAGGTCTT CATGGGATGA AGCTCTTGTC AACCTAAATA CAAACAGAGA      9500
GAGGTTCTCT GAAAGAAGAG GATAATTACT TGGGAGTAGA ATATTGCAAT
                           A
GGGAATCTGC TTGCCGTTAT AAACTATGTG CAAATTCAGG GAGGTAAACA      9600
AGACAAAGAT GCTCCATAGA AAATATGAGA AGAATCTCAT AACTGTTTTG
AGATAATTAT TGTTAGCTAC AAAGATCAAT AACAAGGGTG ATGCCACACC      9700
AAGGTTGGAC AGGCAGTTGC TGGACAGGTG TCCTTGCAGA AATATTTTTG
TGTAAAGTTG AAATAGCCTT TGTGCAAAGT TGTGGTTTTT GTAGACACTT      9800
```

FIGURE 1D

```
TTGTAATAGT TTTGTTTCCA GGAACACAAG CATAAGAATC CTCTCTTCAT
AGCCTTCTTG GGATTTATTT GTCAGGGTTA AAAAACAATT AGTGACATCA      9900
CTTTGGTTCT GATAAAGTTC ACACTCGCTA TTGTAAAACT TTTCGAGGCT
TGTCCTACCA AGGATCCCAT GTGTCACCAG GTATCGAGGT CTTCAGTCTG     10000
AACTAGGCTA GGAGCATTGT GGTTACCACT TTTCTGCAGG TTTTGGTGGC
CCAGGGACTC CCAGCATCGC CTTCTGTCCA GTGTCTGCCT ATTCCCCTCT     10100
TCTTTTTTTC TTCCTTAGGT GCCCTTTTAT CACATGCATT GTCTCAGACC
CTTCTAATAT GTGCTCATAA ATGCATGGCA TCATCTCCTT CCCACATTGA     10200
TTCACTTTCA ATTAAAAGCC AAAACTCCTT CATTTAGACT GAATTTAACA
TGTGCTTTTG AAAGAAGGGT TGAGAGATAA TAGAGAAACA GATTGGGAAA     10300
CCACTTATGC TCCACTTTTT TAAACTTTCT CTGCAAGTAT GGAATTTTTT
GTTCTGCTTT GTTGTTTAAA TTTAAGCCAA AACTTCTTAA TAGAAGGATA     10400
TACAAATATT TATTGGTTTA TACCATTGCA CTTACTTTGA AGAAGAGATG
CTGAATATTA TTAAACCATT GTGTTCCCTG GTGGGCTGAT GGACTGTGAT     10500
TTTATAAGGT GGTCTCAGCC AATTGCAGCA GCTGTTCCCT GTCAGAGGGG
CTAGAGGTTT GGTGAGAGCA GTGGATGAGG TGCAGTGGTG TGTTTGTTCA     10600
CTAGAAGCAA GTGGGAGAAA GCTTTGCCTC TTTGTACTTC TTCATCTTCT
CCCCTCAAGT CCTCAGAATC CACAGCGCTG ACTGTGGAGT GCTGTGGAGC     10700
TGGCATGGCC CATACAGGCA ACATGACTTA GTAGACAGAT GACACAGCTC
TAGATGTCCA TGGGCCCCAC ACCAACTGCC CTTGCAGCAT TTAGTCCTTG     10800
TGAGCACTTG ATGATTTACC TGCCTTCAAT TTTTCACTGA CCTAATATTC
TTTTTGATAA TGAAGTATTT TAAACATATA AAACATTATG GAGAGTGGCA     10900
TAGGAGATAC CCACGTATGT ACCACCCAGC TTAACGAATG CTCTACTGTC
ATTTCTAACC ATAATCTCTT TAAAGAGCTC TTTTGTCTTT CAGTATCTCT     11000
TCCCTGTTTG GACCACATTA CCCTTCATCA TATGAAGCCT TGGGTGGCTC
CTGTGTGAGA CTCTTGCTGT GTGTCACACC CTAATGAACT AGAACCTAAG     11100
GTTGCTGTGT GTCGTACAAC TAGGGGTATG GATTACATAA CATAATGATC
AAAGTCTGGC TTCCTGGGTG TGGCTCCAGC TGCAGAATCG GGCTAGTGAA     11200
                                           A
GTTTAATCAG CTCCGTTGTC CCCACACAGA ACGTATGAAG GTCAACTCCC
            T
     [exon 4: 11230..
TGTGCTGGCC ATCACAGATC CCGACGTGAT CAGAACAGTG CTAGTGAAAG     11300
AATGTTATTC TGTCTTCACA AATCGAAGGG TAAGCATCCA TTTTTTGAAA
        A
                ..11328]
TTTAAATAAT GATTGATCCA CTGATTAAAT TTTTATTTTG AAAAAAACAT     11400
ATATTCACAG AAGGTTACCT AAAAAATGTA CAGGAAGGTT CCATGTACTC
TTCATCCTGT CCCGCCCAGT GGTAACATCT TGCAATCTTG TATATTGCAA     11500
TATATATCTA GTATATTCAT ATTATCAGGT TGGCACAAAA GTTAAAATGG
CAAACTACAG GCTGGGCATA ATGGCTCATG CCTGTAATCC CAGCACTTTG     11600
GGAGGCCGAG GCAGGTGGAT CACGAGGTCA GGAGTTCGAG ATCAGCCTGA
CCAACATGGT GAAACCCCAT CTCTACTAAA AATACAAAAA TTAGCTGCGT     11700
GTGGTGGCAT GCGCCTGTAG TCCCAGCTAC TCAGTAGTCT GAGACAGGAG
AATCGCTTGA ACCTGGGAGG CGGAGGTTGC AGTGAGCCGA GATCACGCCA     11800
TTATACTCCA GTCTGGGCAA CCCAATGAGA CTCCATCTCA AACAACAACA
ACAACAACAA CAACAAAAAC CGGCAAACTG CAATAACTTT TGCACCAACC     11900
TAATACTATA GTACAGGAAA TTGACTTTGA TATAGTTTAC AGAGCTTTTC
AGATTTCACC AGTTTTACAT GCCCTTGTTT GTGTGTGTTT ATGTGTGTGG     12000
GTAGTTCTAA GCAATTTTTC ACATTCGTAG ATTTGTGCAA CGACCAGCAC
CATCAAGATG CAGACCCATT CCGTCACCAT GTGGCTCCCT CCTGCTGTCC     12100
TACAGTCACA ACATGGAGTT TGTCTTTTTC TCTGACAGGT TCTATATCAG
AGCAAACTTT TATTTATTTG AGGAGGCCAA TGTATTAATA TTTCCTTTTA     12200
```

FIGURE 1E

TGGATTGTTC TTTTGGTGTT AAGTCTGAAA ATCCTTTGCT TAGCCCTCCT
TCCTACATTG CTTTTTCTAA GAGTTATATA GTTTAACACT TTACAAAATG 12300
TAACTCTATT ACCCATTTTG TGTTAATATT TGCATAAGTT ATGAGATTTA
GATCAAGGTT CATTTTCTGT GGACTATGGC TGTCCAAATG TTCCAACACC 12400
ATTTTGGAAA GGTAGGCATA TTGTCAAAAC TCAGCTGAGT ATATTTTGTG
AATCTATTTC TTATTGTTTA CTCCTCCACT AATACCACAC TGTGGTGACT 12500
CTAGTAGCTG TACAGTAACT CTTAACATCA TATAGGGCAA TTCTTTCCAC
TTTATTGATT TATATTTTCA GAATGGCTTT AGCTTTTCTT GTCCCTTGCC 12600
TTTCCATAAA AATTCAGAAT AAGCTTGTAA GTGTCTACAA ACAAACCTGC
CATAATTTTG ATAAGAATTA AAGCAGAGGT GTCCAATCTT TTGGCTTCCC 12700
TGGGCCACAG TGGAAGAAGA AGTGTCGTGG GCCACACATA AAATACACAC
ACACACACAC ACACACACAC ACACACACAC ACACACACAC AAATGGTCTG 12800
TGTATAGTTT TCATTATATA TCTACCACCA CAGATAAGCA AAAATGTCCT
TGCATAATAA TCCTAATTAT GCACTGCCCC ATTCAGAGGG TCTTTCAAAA 12900
TCATTGAACA GGTTCCAAGT TTGCAATCAC TGATACAGAA AATGTACATA
TCTAGCTAAA CTTCACTACT TTTTTGATAT TTTTTATTAT AAAAGAAAAG 13000
AGAACAACAT AAAACTAGTG GGGTACTTGA CATTGTTTTT GAGAAACTAA
TCCATCAGTA TCTGGCTTGA TGGAAGTAGT TGCAATTCTC AGTGAGTTCT 13100
CAAGGTGCTC ATCAGATATT TTGGTTCTAA TTTTACTCTT CGTGTTCTTC
ATCCTTGAAA ATAGTAGCTC ACAAATGTAA GTGCTGCCAA AAAGCAATGA 13200
CATGAACAAG GTGTGATTGT GAAGCAAGGG ATATTTGTCA TTGGGAAGAC
AGGTCTTACA AAAGTCCAGT AAAGAGGCAA AATCAAATTT TTCTATAAGT 13300
TGAACATCAG ATTGCAGCTC TAGGCATTCC ATTTCAAAAT TGCCAGGTAA
CATATATATG TCGACTGAAA ATGGAGTTGC AAATATACCA AAATATTGAT 13400
GATTTTTTCA GAAATCTTGA AATACCTGTT TTCAAATTCC TGTATCAAAT
TGAAAAGCAA GGCTGCGTAT TTTTGGCTGT TCACAGGACC ATGTTTAGCC 13500
AACATGTCGA AATGCATAAA ATTGTTTGCC TTAATTTGAG CTTGCCATAA
TTTCAGTTTC ATATGGAATG CTGTTATGGT TTGAAACATT GTATTGTTAA 13600
GTTGGTTTTC AACTTGAAGA CACAGGTTTA ACTCACTTAA ATGGGCCGTC
AAACCCACTA AAAATGCTAA ATCTGTAAGC CAGTTTTCAT TGTCAAGTTC 13700
TGGCACCAAT TTTGTTTGAT ACCATAAACA GCTTGATTTC ACATCACAAA
GCATAAAATC TTTACATTTT GCCTTGACTT AACCATCTTA CTTCTAAAAA 13800
GTGAATGACT TGCTAGAGTC AGCATCCATA CTTTTAAGGA ATTCCTGAAA
CTAGCGATGA TTCAATTCCT GGGCCCTTGT GAAATTTACA GCCTTGATGA 13900
CAATTTGCAT GACGTTATCT ACTTTTAAAG CTTGTGCACA TGGATTTTCT
TGATGTATTA TGCAATAATA CTTCATCAAA TGTGAGTTTT GTGTGGCAAC 14000
TGCATCATCT ATTAATTGTA CAAGTCCCTC TCTTTTACCT ACCATCGCCA
GGGCAGCATC TGTAGCTATA TCACATATGT TTACAAAGGA CAAAGAAAAT 14100
TGCTTTAACA TATTTTTCAC TGCTTCATAT AAATCTCTTG ATTTAGTTGT
GTCTTTTAAT AGCATGGTGA CATTTCGATT TCTTCAGTGA CATTATATTC 14200
ATCATCAATA CCTCTAATAA AAATAGCAAG TTGTGCCGTA TCTGTAGTGT
CAGTGCCTTC ATCCATCACC AAAGCATAAA ATTTTAAATT AGCAGTTTTA 14300
CTCTCCAAAC GTCTTTCAAT AGATTTCCCA ATTTCTCCAA TTCTCCTGGC
TATAGTCTGG TGAGACAAAC TGATTTTAGA AATATCAGTT TCTCAAGGCA 14400
AATAATATCT ACCACATCTT CCAGACATTG CTTAATAAAC TCACCATCAG
TAAATGGTTT TGATTTTTTT GCTATTAAAT TTGCTACCAC ATAACTAGGT 14500
TTTACCTTAC GATTGAGTCC GAGTTGTAAC TTTTTAAAAA ATCTTTTTTG
TTGAAAAGAC AGACTTTTTT TCAGTTCTGC TATTTTGTCC TTACAACACA 14600
TACACACCAA ATTTGTCAGC ACGTTTTTGC ATATAATGCC TCTTCAAATT
GTAGTCTTTG AAAACTGGCA CAAATTCCGT GGAAATTAAG CAGAGTGCTT 14700
CGCTATTTGC CTCAACAAGA AAAAGTCATT TGTCCACTTT TCATTGAACA
ATCTTCCTTC ATCCATAATT TTTGTTTTTT AGGGTTTTCT TTTTAAGACA 14800
TTGTGGAAGC CATTCTGGAA TTAAAAGCAT TATAATAGAT AAGCAACTAT

FIGURE 1F

```
ATTTACTTTT ATTATGGAAA TTAACAGATA GGAAAATAGA ACAGAAAGCA      14900
AGGTTTAATA ATCAAATAAG AATACTTACA TGTCTTCTAA ATAATATTAA
ACACCTATCA TCTACAAAGG TAGGTTGAAA TATTATTGAT AATTGCTGGG      15000
TTTTACTTGC CAAATTGCCA CAAACACACC TAATACCTGA CAGTGTCAAT
TCAACTGTCC GTGATTAGAA GATAACACAC TGGAAGTCGC ACACCACCAT      15100
AAAACTGAAG CCACACATGC GTACAAATGG CGACAGTGTC TGGTGTACAG
CAGCGCTCTG CCTTGTCCAG AATACACACT TGAATTCTTT GTCACAATTC      15200
ACTTCACGTG GCACTGCAAT AGCGTCCTCT CGCTCTTTGT TAGTTAATTT
TAATGGCTTT TAATTTCTTC TTGCTGAACT GTTTGCAATT ATAATGCAAA      15300
TTATGGATAC TAGTCCATTA TTTGTGGATG TGACATACTC TGATTACCCC
TTTCCATTCC ATTGTTGTCT ACGAAGTTCA CACTTGAGAA TCACATAGTC      15400
AAATTACAAA ATTACAAAAA AAATTGCAAA AAAACTCAAA ATGTTTTAAG
AAAGTTTCCA CATTTGTATT GGGATACATT CAAAGCCATC CTGGACTGCA      15500
TGAGGCCTGC AGGCCACAAG TTGGACAAGC TTGAATTAAA CCAATAGAAC
AATTTGGGTA TAATCTATAT CTTTACTATG TTCAGCCTTT CATCCCGTGA      15600
ATATAGTATG CCTCTCCATT TCTTTAGCTT TTATTACTTT CCTCAACATT
TTATAGTTTT CAGCATAGAG GTCCTGTACA TCTTTTGTTA GATTTACACC      15700
AGAAATATTT CATTTTTGTT GGAGTAACTG TAAATGATAC TGTTTTTCTT
GTATTTTCAG ATATTGATTA TTGTTACATA GAAATGTGAA TAATTTTGTT      15800
TGTTGATCTT GTATCCTATA GCCTTGCAGA ACTTACCTAT TCGTTCTAGA
AATTTTTTTG TATATTCCTT GACATTTTAT ACATTGACAA TTATGTCACC      15900
TGAAAATAGA GACAATTCTA TTATTTCCTT TCCAATCTGT ATGCCTTTTA
TTTCTTTTTC TTGTCTAGTG TATTAAGACA TCAGGTATGC TCTTTAGTAA      16000
GAATGTTGAG AGTGGGCATT TTTTAGTTCT TCTTGATCTT GGAAAAACCA
TTCAGTCCTT CATCATTAAA TGTGATTTAA CTGAATGATT TTTTTACAGA      16100
TTGTCTTTAT CAAATGAAGG AACTGTCTCT CTCTTCCTAG TTTATTGAGA
TTTTATCATG ACAGCTGGAA GTACACATTT TAAAACAAAA CATAGTTGTG      16200
GAAGATAAGA GAAAGTTCCA AGCATGCTGG CTTGATAGTC CAGCCCCAAG
TTGGGAAAAG TAATTATCCC TTTCTTTTTC CTTCTATTTA TGGAATAAAA      16300
AATTAAGAGA AAAGAATTTT CAAGGAAATT GCATTATTCC TTCAAAACAG
GTTTCTAGTC TTTAAGTATT ACCTACTTTT CAAAAAAAAA TCACCACATC      16400
ATGGCATCCC TTTTTCAAGT TGCCCATGCT GTAGGTGTAT TAAAGACAGA
GCTGGTCTGA GGCAACATAC AGTCTGCCCA TCTGTCACCA ATCCTTTTCT      16500
ACTCTGCACA CTCCTGGGGA AGGGCTAGGT CTTGTTCCTG TCTATTCCAC
TGGAAGAACA GTTCCCTACC ACGTGGAGCA TTTGCAATTA AAAGGAGACT      16600
GAGATATAGA GGCAGGAGAC CACACCAGAT GGCTGGGTCT CCCCACTCCC
ACCCCCGCCC CACATACACT CAGAAGAGGC TAGGCATCTA GGATCTCCAT      16700
TGAGCATCTT GAATATGGCT TGCCATAATA TCATATACAG TCAATAAATA
TTTGTTAAAT AAGGATGCCT CTTCAATATA TTTTGTGCAA CCATGAAGAT      16800
CACCACAACT AATGTGAGAA AAAATGTTTC TGTTGAACTC TAGTCTTTAG
                                  T
     [exon 5: 16843..
GCCCAGTGGG ATTTATGAAA AGTGCCATCT CTTTAGCTGA GGATGAAGAA      16900
TGGAAGAGAA TACGGTCATT GCTGTCTCCA ACCTTCACCA GCGGAAAACT
CAAGGAGGTA TGAAAATAAG ATGAGTCTTA ATTAGAAATG TAAAGAATGA      17000
              ..16957]
ATCTGGGGAC AGGTAGAAAG TAAGATCACA GTCCGTTTCC AAGGGGTAGT
CCACTGAGTT CGAGCTTCCT AAAAATGGTC TTTTATCTTT ATGTACAGAA      17100
AAGACATCAC AAAATTCATT ACAAAATGTC ACTTACTGCT CCATGCTGGA
GAAAGCCATA TCCTTCTGGG ACTTGAGTCT GCACATTTAA CTACAGGTAC      17200
TGATCTGTTT TGTGCTTAGA TGTTCCCCAT CATTGCCCAG TATGGAGATG
     [exon 6: 17220..
TATTGGTGAG AAACTTGAGG CGGGAAGCAG AGAAAGGCAA GCCTGTCACC      17300
```

FIGURE 1G

```
TTGAAAGAGT AAGTAGGAGC ACAGCCATGG GGTTCTGAGC TGTCATGAGC
                ..17308]
CCTTCCAGCT GCCTGCCATG GAGTCGACAG TCGCACTGTT GGGTTACTCC      17400
                                    A
AGTGACCAGA CAAAAGCAGG GCAGCGCTGC AACTCCAAAG AGCCACCTAA
GAGGGAGTGG CTCCCATGAG GCGGCAAGTC AGCAAGGGAA AAGGGCCTTC      17500
TCTCCTGTGC ACAGGAGCCA GGATTTACTT ATCTGTTAAC TTGTCACCAT
AAATATTCTG GGAGATTAAA TACATACTTT AGAAATTAAA AAAACATGAT      17600
TGTATCAAAG TTTTGAGTGT AGTGGATATG GAACTGTGGG TAAGCAAGCA
TTTGGTACTT GTTGCCTTGC ATTGGGTAAG ATGGGAAAGT TACAATGGGG      17700
AACTTGGAAC AATTTCAATC CCTTCATGGT TTTTCTGAGA ATATCAGCAA
ACTATGAACT ATTAAACCTT CCCACTACTT CCTTTTCCTC CAATCTCAAA      17800
AAAGAAAGGG TGCTAGAAAT GCTATGTGTA GAGCAAGCCT ATTATTTGCT
GTCTACAATG GTATGTGCTT CAATTATGCA GGAACGACAG GTGTAATCTG      17900
AGCCTGTCCT GTTCAGACTT GGGACATGTG GTCACTCAGT TTTGGGTTCT
CCAAATCAAT GTTGGAGAGA TCTATTTTTT TTAACCAGAA CATTCTTGAT      18000
TGTCACATCT TACAAAAATG ACTCTGCTCT CAGCGCAACT TCAGGTCAGA
GGAGCTGGGG ATAGTGGGGT TTTCCAGAGC ATTAGCAGGG AGTGTAGAGA      18100
ATAAAGGATG ATATTTCTAG GAACTCAGAA CAGGGTGTTA CTGTTTTGTA
AAGTGTTGAA GAGGAATTGG CTCTGGGCAT AGAGTCTGTA GTCAGACAAC      18200
GCCACCTTTC TTGAATCCAC TAGGAAGAGT TAATTATTCT ACTCTTGTTC
TGCTGAAGCA CAGAGCTTAC ATATCTTATA TCATCCACAC TCAACACATG      18300
CTACTGTAGT TGTCTGATAA TGGGTCTCTG TCTTCCTATG ACTGGGCTCC
TTGACCTCAG AGGTGAGTCT AACTCAGCTT GGTGTCTCCA TCACCCCCAG      18400
CATAGGGCCA GCTCCATCAC TGGCACCAGA TAACCACCTT CTGAGGGAGT
AGATGGAAGA TGATTCAGCA GATAGTTCTG AAAGTCTGTG GCTCTTTATG      18500
TGTCTTGACT GGATATGTGG GTTTCTTGCT GCATGTATAG TGGAAGGACG
GTAAGAGGTG CTGATTTTAA TTTTCCATAT CTTTCTCCAC TCAGCATCTT      18600
     [exon 7: 18595..
TGGGGCCTAC AGCATGGATG TGATTACTGG CACATCATTT GGAGTGAACA
TCGACTCTCT CAACAATCCA CAAGACCCCT TTGTGGAGAG CACTAAGAAG      18700
                                                    A
TTCCTAAAAT TTGGTTTCTT AGATCCATTA TTTCTCTCAA TAAGTATGTG
                               G
                ..18743]
GGCTATTATT TCTTTCTCTC TTTTTAAAAA TAACTGCTTT CTTGACATAT      18800
                                    T
AATTCACATA TCGTATAATT CATCCACTTA AAAGGTACAA TTCCATTGTT
TTTAAGATAA TCAAAAATAT GTATGACCAT TACTATTGTA AACTAAAATG      18900
TTTTTGTCAA TCTAGAGCCC TCACACACTT TAGCTGTCAA CACCCCACCA
CAAACCCCAC TGCCCTAAGC ATCCAATAAT CAACTTTCTG CCTCTATAGA      19000
TTTGCCTATT CTGGACACTT CATAGAAATA ATATCATTGA TTTTTCTCTG
TTGTTTTTTA TTCTCTATTT CATGAGTTTA TTTTAGTCTG TTATTTTCTT      19100
TCTTTTGCTG GCTTTAGGTT TCATTTGCTC TTCTTCTTTT AGTGTTTTGT
GGTGTAAATA ATTATAATCA ATTTGAGATA TTTTCTTCTT TTAAATTTAG      19200
ATATTACAGC TATAAATTTC CCTCTGAGCA CTGGTTTGGC TACATCCTGT
GTTTTGGTAC ATCATGCCTT CTTTTTGTTC ATCTCAAAAC AATTTCTTGT      19300
TGCCCTTTTG ATTTCTGCTT TGACTCACTG GTCACTTAAA ACTGTATTGT
TTAACTTCCA CAAATGTATG AGTTTCCCAA ATTTCTTTCC CTTATTGATT      19400
TCTAGTTTTA TTCCATGGAA GTTGATGTAC ATATGCTGTG TTAATTCTAT
CTTGACTATC ATTTCCTGAA CAGCATGATT AAGTTAAGCA GCAGATTATG      19500
GTCTACATTA ATCCAAAAAC TCTAGTCCAA TAGATAAAGG CTAAGAGGTC
AGGGAATTTA ATTCTATTAC TTTGGTCACT CCAAAGACTC AGAAGGTGCC      19600
```

FIGURE 1H

```
ATTGATCTCA CTGCTGTAGT GGTGTTTCCT ATGTATAGAC CTGCCCTTGC
TCAGTCGCCG GCCTGAAAGA AGGGCAAACA TGATAAAAGG AATGGGTTCC      19700
AGTTGAGAAT CATGATGTTC TTATTCTTAT TACTGGTAGA GAAAATTATA
ATTGCTCCAG GTAAAGTTTG CATTTTCAAT GATTTCCTTT TGTTTGTTTT      19800
    T
GTTTTTCCCA CAGTACTCTT TCCATTCCTT ACCCCAGTTT TTGAAGCATT
     C
      [exon 8: 19814..
AAATGTCTCT CTGTTTCCAA AAGATACCAT AAATTTTTTA AGTAAATCTG      19900
TAAACAGAAT GAAGAAAAGT CGCCTCAACG ACAAACAAAA GGTAAAATCT
                 ..19941]
GATGGTGGTT AAATGACGAT GTTTAGGTTT TGATAAATTT AGATTTTATA      20000
CACATGATAG AGCATGTATC TGTATTTTTA AAAATAAAGA CAGAGAACTT
ATGTTTAGAA CAAGAGAAGC CATTTGGTAG AAATAAAGAA GGAGATTGGG      20100
               C
GAAGGAGATG AGAATGAGTC AGAGAGATAG CATTTAAAAC TTGAAATCAG
GCACAACAAT TAGTATGTCA TGATATAAAC AGTATTGAGA TAAAATTTTA      20200
CCACTTCTCT TCCCTTTAAT AAATTGTCAA AGGATAAAGT TTCCTGTTTG
AAAATATATT TTACTGGTAT TGTGCTTTCC TCATATCACA GATTGGTAAA      20300
GAATCATTTT AAGTCCAAGA CTCTTATTTT ACATATTCTG CAATTAAAGG
TCCTATGAGG CTACCTGCCG ACTGCTGACA TGTAGTGTGT GGTAAATGTG      20400
AGTGTTTCAC AGCCTGGAGT GAACAGGGGT CTTCTCTGAG AATTGAGGTT
GCAAGGCTGG CTAACTCAGC TTTGCCTTCA CGAGCCCTAG AGGCCAGCCG      20500
AAGGATGTCT GCAGGTCAGG GAGACAGGAC CAGGTAACCC AGCTGTCACT
GAAGATTATA TAGAGTTTGA GAATGTTGGA ATATTTGAAA ATGCTCCCCC      20600
AAAAAAGCTG CTGATGAGTT CTGGAAATGT CAGGAGATTA ATCTATACGG
ACACTGCTGA AGAAAAAGGT AGAAGAATAA AAGATCCAGT ACTTCTTCCT      20700
GGGTAAGCAG TTATGACCAG AGATGGAACC GGCAACTCTT TGGCCAGAAA
GCTGTATCCA AAAGACAGAG AAGATGAGAA ACAGGGAGGG CAAAGGCGAA      20800
AAAGCAATTG GACATGATAG CTAGATTTGT TTCAGGAAAA CATCCTGCTT
TCCAAGGATT TAGATGAATG TTTTTGTTCA CTGGTGACTC AGGTAACACG      20900
TCTTCAAGAA GCCATAGGGA GGTTGAGGGA GGGAAGTCAA GAAGGGAGGT
TGAGGACTGC ACTTTTGATT TACTTCTGAC TTCACGAGTC ACTTTCTGCC      21000
AAAGAAATCT CTCCTTTTGC TTCTAGCACC GACTAGATTT CCTTCAGCTG
      [exon 9: 21027..
ATGATTGACT CCCAGAATTC GAAAGAAACT GAGTCCCACA AAGGTAACCA      21100
                 ..21093]
AGGAGTGCTT CTGAGGGCTA CTGGCGGGGA CACTAAGAGG GAGGGCCTTG
TTCTGAAAAT GTGCAGGAAG TATTCCAGGA AGATGAGAAT TTTTGCCACA      21200
                  T
TAGCAGAACA ACACACATTT AGATGTTATA AATGGTAGCT GGAGGCACTT
TCCAGAAGCC CACAGGTATA GCCATGTTCC AGGCTGAAAG GGCAACCCTA      21300
AGCAAACCTA GAATGCTTGG AGGACAGTCA GTGGTTTGTG GATCACCTAC
ATGAGATCAA ATGCCAGTTC TCAGCCTCCT CCAGATCCAC CAAGTGAGAA      21400
CCTCTACTTG GAAATTTATA TCAAACATAC CGATCAGGAA GCACACTATC
CCAGTAAGGG TGATTTTAAC TGGCAGTACT TGAAAGTGTG TTCGCAAGGT      21500
TAATCTACTG CAAAGTTTTA TTTTTCCCTT TGAAATGCAT AAGTAACTAA
TGGGGGACAC CTCTGATACC ATGTAAATCT ACTTCAATCT TCAGTCTTGT      21600
ATCTACTAGT TTTATGACCC ATGGATGGTT TTAACCAAAA CCATTATTAC
TAAGACAGTG GCAAAATGAT AACCATGGTC AATTTCAAGC TACCAAGATT      21700
TGGCAACCAT CTCACAAAAT TTTTGAATAT TTAACAATTG GTTCTAGAGA
GCAGGACTCA GCAGACTCCA GTATACCACT TTAAACATGT CCATGTCTAC      21800
ATCTACTTCT GTCTGTCTAT CTATCTGTCA ATCATCTATC TGCCTATAAT
```

FIGURE 1I

```
TTATCAATTA ATCATCTATC TATCTCAACA AAACTTGCTG TGATAAAGAA      21900
AATAGTCTAT CATTTCACTG TTTCATATAG AAATCACTAG ACACATATGG
CTATTGAGTA CTGGACATGT GGCCAATGCC ACTGAAGAAC AATTTTTAAG      22000
AGTATTTATT TTTAATTGAA TAAAATTTGA ATTTAAATAG CCACATGTGG
ATAGTGGCTA CCAGATTGGA CAGCAGAGCT CCCAACTTTA AAATTACAGT      22100
TCAATTTCAA CTCAGTATAA TGGGGTTCAA TGTAACTGAG TAAAATAATT
GGATGGTTGA ATTTACCCAC AGCAGCATAC AGAAATATTC ACTGATAAAT      22200
CAGAACTCTG TAGACCTTTC TCACACTCAT TTTATATTGT GTTTGGTTGT
GAGTTACATG ATTGCTGCAG GCACCATATT TATTTCTGTG CTCCAGGTCT      22300
CTAAAGGTCC TAATCCAGTC CTGACCAAAC AGACTAGTGA TGGACCATCG
TGAGCTTCTC TCAGGAGAAA TATCAAGAGG GAGGCCAACC TGTAATCATA      22400
AGAACTTCTG CTATTTTAAT GCCATTCATC AGACTACAGT CAATCACCAT
GCTTCTGGCT TTTTGTCTAT CTCTGCTGTC TTGTACATCC TGAGATAGTC      22500
CATTCTGAGA ACTGTACCCT AGATCTTGTA TTGCCTGATG CCTGTCAAAG
ATGTAATCCA TGCTGCTTAA GTGAGGTTGT GCACACAAAT CACCATATCT      22600
CCTGCAAGTT TGGATTTTGA TTCAGTAGTT CGATGGTGGG GTTTGAGATT
CTGCATTTCT AATAAGCTCC CAGATGTGGC TGGTGCTGCT GGTCCATGAA      22700
ACACACTTTG AGTAGCAAGA GGTGATCTGT AGCTCAGTAT TGGTCCTTTA
AGTTCCCTCA AACATATATA GAGAAAAGGT CCTAAATATT GCAAATTCTC      22800
TCAAAGTTTG TCAAGCTATA TTGGAATTCT CTCAAAGTCT GTCAAGCTCT
ATTGTAGAAA ATCAAATTTT TATTGGGAAA AAGCCTACCC CATATTTACT      22900
TACAGATAAA GTACTTTTAG GATCATTCAA GGCACACACC CATAACACTG
AGTATGTAAG ACAGAAATGC TCTCTCTGGA AATTACAGCA GTGCTGGTGC      23000
TGGGATGCCA TGATGAGGAG TGTGTGGCCC ACAATCATGT AGACCTTGGG
AAAACCTGGA TTAAAATGAT TTTGCGTCAT CCTGGCCCTG TATAAGATAC      23100
ATATCAGAAT GAAAACCACT CCCAGTGTGA CTTTGAATTG CTTTTCCATT
TTTTCTTCTT GGGATTAGAG AGCTTCACTT AGATTTCATC TAAGCTGTGA      23200
TGTTGTACGT TGACCTGATT TACCTAAAAT GTCTTTCCTC TCCTTTCAGC
     [exon 10: 23250..
TCTGTCTGAT CTGGAGCTCG CAGCCCAGTC AATAATCTTC ATTTTTGCTG      23300
GCTATGAAAC CACCAGCAGT GTTCTTTCCT TCACTTTATA TGAACTGGCC
ACTCACCCTG ATGTCCAGCA GAAACTGCAA AAGGAGATTG ATGCAGTTTT      23400
GCCCAATAAG GTGAGGGGAT GACCCCTGGA GATGAAGGGA AGAGGTGAAG
                ..23410]
CCTTAGCAAA AATGCCTCCT CACCACTCCC CAGGAGAATT TTTATAAAAA      23500
GCATAATCAC TGATTCCTTC ACTGACATAA TGTAGGAAGC CTCTGAGGAG
AAAAACAAAG GGAGAAACAT AGAGAACGGT TGCTACTGGC AGAAGCATAA      23600
GATCTTTGTA CAATATTGCT GGCCCTGGTT CACCTGTTTA CTGTTATCAC
AATAATGCTA AGTAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGAGTG      23700
TGGCGAGAAG ATGGCCAAAC AGGAACAGCT CCAGTCTACA GCTCCCAGCG
TGAGCAACAC AGAAGACGAA TGATTTCTGC ATTTCCAACT GAGGTACCGG      23800
GTGCATCTCA ATGGGGATTG TTGGAGAGTG GGTGCAGGAC AGTGGGTGCA
GTGCACCCAG CCTGAGCCAA AGCAGGGCGA GGCATCACCT CACCTGGGAA      23900
GTGCAAGGGG TCAGGGAATT CCCTTTCCTA GGGGTGACGG ACAGCACCTG
GAAAATCAGG TCACTCCCAC CCTAATACTG CGCTTTTCTG ATGGTCTTAG      24000
CAAACGGCAC ACCAGGAGAT TATATCCCGC GCATGGCTCG GAGGGTCCTA
CGCCCATGGA GCCTCGCTCA TTGCTAGCAC AGCAGTCTGA GATCGAACTG      24100
CAAGGCAGCA GCAAGGCTGG GGGAGGGGCG CCCGCCATTG CTAAGGCTTG
AGTAGGTAAA CAAAGCTGCC AGGAAGCTCA AACTGGGTGA AGCCCACCGC      24200
AGCTCAAGGA GGTCTGCCTG CCTCTGTAGA CTCCACCTCT AGGGGCAGAG
CATAGCCAAC CAAAAGGCAG CAGAAACCTC TGCAGACTTA AATGTCCCTG      24300
TCTGACAGCT TTGAAGAGAG TAGTGGTTCT CCCAGCACAC AGCTGGAGAT
CTGAGAACAG ACAGACTGCC TCCTCAAGTG GGTCCCTGAC CCCCGAGCAG      24400
```

FIGURE 1J

CCTAACTGGG AGGCACCCCC CAGTAGGGGC AGACTGACAC CTCACACGGC
CGGGTACTCC TCTGAGACAA AACTTCCAGA GGAATGATCA GGCAGCAGCA 24500
TTTGCGGGTC ACCAATACCG CTGTTCTGCA GCCTCCACTC CTGATACCCA
GGCAAACAGG GTCTGGAGTG GACCTCCGGC AAACTCCAAC AGACCTGCAG 24600
CTGAGGATCC TGACTGTCAG AAGGAAAACT AACAAACAGA AAGGACATCC
ACACCAAAAC CCATCTGTAC ATCACCATCA TCAAAGATCA AAGGTAGATA 24700
AAAACACAAA GATGGGGGAA AAACAGCAGA AAAACTGAAA AATCTAAAAA
TCAGAGCACC TCTCCTCCTC CAAAGGAACG CAGCTCCGCA CCAGCAACGG 24800
AAAGCTGGAT GGAGAATGAC TTTGACGAGT TGAGAGAAGA AGGCTTCAGA
CGATCAAACT ACTCCGAGCT AAAGGAGGAA GTTCGAACCC ATGGCAAAGA 24900
AGTTAAAAAC CTTGAAAAAA GATTAGACAA ATGGCTAACT AGAATAATCA
ATGCAGAGAA GTCCTTAAAG GACCTGATGG AGCTGAAGAC CATGGCACGA 25000
GAACTACGTG ATGAATGCAC AAGCCTCAGT AGCCAATTCA ATCAACTGGA
AGAAAGGGTA TCAGTGATGG AAGATCAAAT GAATGAAATG AAGAAAGAAG 25100
AGAAGTTTAG AAGAAAAAGA ATAAAAAGAA AGGAACAAAG CCTCCAAGAA
ATATGGGACT ATGTGAAAAG ACCAAATCTA CGTCTGATTG GTGTACCTGA 25200
AAGTGACGGG GAGAATAGAA CGAAGTTGGA AAACACTCTG CAGGATATTA
TCCAGGAGAA CTTCCCCAAT CTAGCAAGGC AGGCCAACAT TCAAATTCAG 25300
GAAATACAGA GAACGCCACA AAGATACTCC TCGAGAAGAG CAACTCCAAG
ACACATAATT GTCAGATTCA CCAAAGTTGA AATGAAGGAA AAAATGTTAA 25400
GGGCAGCCAG AGAGAAAGGT CGGGTTACCC ACAAACACAA ACCCATCAGA
CTAACAGTGG ATCTCTCGGC AGAAACTCTA CAAGCCAGTA GAGAGTGGGG 25500
GCCAATATTC AACATTCTTA AAGAAAAGAA TTTTCAACCC AGAATTTCAT
TTCCAGCCAA ACTAAGCTTC ATAAGTGAAG GAGAAATAAA ATACTTTACA 25600
GACAAGCAAA TGCTGAGAGA TTTTGTCACC ACCAGGCCTG CCCTAAAAGA
GCTCTTGAAG GAAGCACTAA ACATGGAAAG GAACAACTGG TACCAGCCAC 25700
TGCAAAAACA TGCCAAATTG TAAAGACCAT CGAGGCTAAG GAGAAACTGC
ATCAACTAAC GAGCAAAATA ATCAGCTAAC ATCATAATGA CAGGATCAAA 25800
TTCACATATA AAAATATTAA CCTTAAATGT AAACGGGCTA AATGCTCCAA
TTAAAAGACA CAGACTGGCA AACTGGATAG AGTCAAGACC CATCGGTGTG 25900
CTGTATTCAG GAAACCCATC TCACGTGCAA AGTAACACAT AGGCTCAAAA
TAAAGGGATG GAGGAAGATC TACCAAGCAA ATGGACAACA AAAAAAGGCA 26000
GGGGTTGCAA TCCTACTCTC TGATAAAACA GGCTTTAAAC CAACAAAGAT
CAAAAGAGAC AAAGAAGGCC ATTACATAAT GGTAAAGGGA TCAATTCAAC 26100
AAGAAGAGCT AACTATCCTA AATATATATG CACCCAATAC AGGAGCACCC
AGATTCATGA AGCAAGTCTT TAGAGACTTA CAAAGAGAGT TAGACTCCCA 26200
CACAATAATA ATGGAAGACT TTAACACCAC ACTGTCAACA CTAGACAGAT
CAACAGGACA GAAAGTTAAG AAGGATATCC AGGAATTGAA CTCAGCTCTG 26300
CACAAAGTGG ACATAATAGA CATCTACAGA ACTCTCCACC CCAAATCAAC
AGAATATACA TTCTTTTCAG CACCACACCA CACCTATTCC AAAATTAACC 26400
ACATAGTTGG AAGTAAAGCA CTCCTCAGCA AATGTAAAAG AACAGACATT
ATAACAAACT GTCTCTCAGA CCACAGTGCA ATCAAACTAG AACTCAGGAT 26500
TCAGAAACTC ACTCAAAACC GCTCAACTAC ATGGAAACTG AACAACCTGC
TCCTGAATGA CTACTGGGTA CATAACGAAA TGAAGGCAGA AATAAAGATG 26600
TTCTTTGAAA CCAACAAGAA CAAAGACACA ACATACCAGA ATCTCTGGGC
CACATTCAAA GCAATGTGTA GAGGGAAATT TATAGCACTA AATGCCTACA 26700
AGAGAAAGCA GGAAAGATCT AACATTGACA CCCTAACATC ACAATGAAAA
GAACTAGAGA AGCAGGAGCA AACACATTCA AAAGATAGCA GAAGGCAAGA 26800
AATAACTAAG ATCAGAGCAG AACTGAAGGA AACAGAGACA CAAAAAAACC
CTTCAAAAAA ATCAATGAAT CCAGGAGCTG GTTTTTTGAA AAGATCAACA 26900
AAATTGATAG AATGCTAGCA AGACTAATAA AGAAGAAAAG AGAGAAGAAT
CAAATAGATG CAATAAAAAT GATAAAGGGG ATATCACCAC CCATCCCACA 27000
GAAATACAAA CTACCATCAG AGAATACTAT AAACACCTCT ATGCAAATAA

FIGURE 1K

ACTAGAAAAT CTAGAAGAAA TGGATAAATT CCTCGACACA TACACTCTCC 27100
CAAGACTAAA CCAGGAAGAA GTTGAAACTC TGAATAGACC AATAACAGGT
TCTGAAATTG AGGCAATAAT TAATAGCTTA CCAACCAAAA AAAGTCCAGG 27200
ACCAGATGGA TTCACCGCCG AATTCTACCA GAGGTACAAG GAGGACCTGG
TACCATTCTT TCTGAAACTA TTCCAATCAA TAGAAAAAGA GGGAATCCTC 27300
CCTAACTCAT TTTATGAGGC CAGCATCATC CTGATACCAA AGCCTGGCAG
AGACACAACC AAAAAAGAGA ATTTTAGACC AATATCCCTG ATGAACAGTG 27400
ATACAAAAAT CCTCAATAAA ATACTGGCAA ACCGAATCCA GCAGCACATC
AAAAAGCTTA TCCACCATGA TCAAGTGGGC TTCATCCCTG GGATGCAAGG 27500
CTGGTTCAAC ATACGCAAAT CAATAAACAT AATCCAGCAT ATAAACAGAA
CCAACGACAA AACCCACATG ATTATCTCAA TAGATGCAGA AAAGGCCTTT 27600
AACAAAATTC AACAGCCCTT CATGCTAAAA ACTCTGAATA AATTAGGTAT
TGATGGAACC TATCTCAAAA TAATAAGAGC AAATTTATGA CAAACCCACA 27700
GCCAATATCA TACTGAATGG ACAAAAACTG GAATCATTCC CTTTGAAAAC
TGGCACAAGA CAGGGATGCC CTCTCTCACC ACTCCTATTC AACATAGTGT 27800
TGGAAGTTCT GGCCAGGGCA ATCAGGCAAG AGAAAGAAAT AAAGGGTATT
CAATTAGGAA AAGAGGAAGT CAAATTGTCC CTGTTTGCAG ATGACATGAT 27900
TGTATATCTA GAAAACCCCA TCGTCTCAGC CCAAAATCTC CTTAAGCTGA
TAAACAACTT CAGCAAAGTA TCAGGATACA AAATCAATGT GCAAAAATCA 28000
CAAATATTCT TATACACCAA TAACAGACAA ACAGAGAGCC AAATCATGAG
TGAACTCCCA TTCACAATTG CTTCAAAGAC AATAAAATAC CTAGGAATTC 28100
AACTTACAAG GGATGTGAAG GACCTCTTCA AGGAGAATTA CAAACCACTG
CTCAATGAAA TAAAAGAAGA TACAAACAAA TGGAACAACA TTCCATGCTC 28200
ATGGGTAGGA AGAATCAATA TCATGAAAAT GGCCATACTG CCCAAGGTAA
TTTATAGATT CAGTGCCATC GCCATCAAGC TACCAATGAC TTTCTTCACA 28300
GAACTGGAAA AAACTACTTT AAAGTTCATA TGGAACCAAA AAAGAGCCCG
CATTGCCAAG TCAATCCTAA GCCAAAAGAA CAAAGCCGGA GGCATCATGC 28400
TACCTGACTT CAAACTATAC TACAAGGCTA CAGTAACCAA AACAGCATGG
TACTGGTACC AAAACAGAGA TATTGATCAA TGGAGCAGAA CAGAGCCCTG 28500
AGAAAGAATG CCACATATCT ACAACCATCT GATCTTTGAC AAACCTGACA
AAAACAAGCA GTGGGGAAAG GATTCCCTAT TTAATAAATG GTGCTGGGAA 28600
AACTGGCTAG CCATATATAG AAAGCTGAAA CTGGATCCCT TCCTTACACC
TTATACAAAA ATTAATTCAA GATGGATTAA AGACTTACAT GTTAGACCTA 28700
AAACCATAAA AACCCTAGAA GAAAACCTAG GCAATATCAT TCAATACAGA
GGCATGGGCA AGGACTTCAT GTCTAAAACA CCAAAAGCAA TGGCAACAAA 28800
AGCCAAAATT GACAAATGGG ATCTAATGAA ACTAAAGAGC TTCTGCACAG
CAAAAGAAAC TACCATCAGA GTGAACAGGC AACCGACAGA ATGGGAGAAA 28900
ATTTTTGCAA CCTACTCATC TGACAAAGGG CTAATATCCA GAATCTACAA
TGATCTCAAA CAAATTTACA AGAAAAAAAC ACAACCCCAT CAACAAGTGG 29000
GGGAAGGATA TGAACAGACA CTTCTCAAAA GACATTTATG CAGCCAATAG
ACACATGAAA AAATGTTCAT CATCACTGGC CATCAAAGAA ATGCAAATCA 29100
AAACCACAAT GAGATACCAT CTCACGCCAG TTAGAATGGC GATCATTAAA
AAGTCAGGAA ACAACAGGTG CTGGAGAGGA TGTGGAGAAA ACAGGAACAC 29200
TTTTACACTG TTGGTGGGAC TGTAAACTAG TTCAACCATT GTGGAAGTCA
GTGTGGTGAT TCCTCAGGGA TCTAGAACTA GAAATACCAT TTGACCCAGC 29300
CATCCCATTA CTGGGTATAT ACCCAAAGGA TTATAAATCA TCCTGCTATA
AACACACATG CACACTTATG TTTATTGCAG CACTATTCAC AATAGCAAAG 29400
ACTTGGAACC AACCCAAATG TCCAATAATG ATAGACTGGA TTAAGAAAAT
GTGGCACATA TACACCATGG AATGCTATGC AGCCATAAAA AATGATGAGT 29500
TCATGTCCTT TGTAGAGACA TGGATGAAGC TGGAAACCAT CATTCTCAGC
AAACTATGGC AAGGACAAAA AACCAAACAC TGTATGTTCT CACTCGTAGG 29600
TGGGAATTGA ACAATGAGAA CACATGGACA CAGGAAGGGG AATATCACAC
ACTGGGGCCT GTTTTGGGGT GGGAGGAGTG GGGAGGGATA GCATTAGGAG 29700

FIGURE 1L

```
ATATACCGAA TGTTAAATGA CGAGTTAATG GGTGCAGCAC ACCAACATGG
CATAGGTATA CATATGTAAC AAACCTGCAC GTTGTGTACA TGTACCCTAA      29800
AACTTAAAGT ATAAAAAAAA AAATTCAAAA ACCTCAGTGG CATCTAATGA
GAAGCATTTA TTGCTCACAA GACTGGATAG TGAGTTCTGC TGATACTGAC      29900
TGGACTCACT CTGGTCTGGC TATGGTCTGA GGTAGCCTGG CCCTGGGGGC
GCGATGGAGG CTGACTCAGC TCTCCCCACA CCTGTCTCAT GTTCCAGTCA      30000
GGTAGCCACT GGCCAAGAAG CCAAGCTAGG AACCAGGGTA TCTGACTCCT
GAGCTAAACT CTAACCCTCT ACAATACTGC CTCCCAAATA TAACACCAAG      30100
TGCTAGGTAC ATATCATCCA CAGTTTTCAG ACTTCTGCCC AAACTGGGAT
TCTTTTTAGT GTGAAGAGAC CTGGCCTGTG GGGCTGACCC TGGTGTGGCT      30200
GTGAGGCAGA CACAAAGGGA CATTTACATC CAGTCCTGAA GATTACAGTC
CAGCCCTGAA GCAACAACTA GGAAACTATT CCAAAAGGAG GGGATGGGGC      30300
TGAGTGTGGG GTTCTATTCT CTTCATAACT TTAACTAGAA CTCAAATTGT
GTACCTTGGT AGCATCCAAT CATAAATTTA TTTTGTCGTA TTTGTGATAG      30400
AAAGGAACAA GTTTATCCAC AAATTTATTT ATTTATTTAT TTATTTATTT
ATTTATTTGA GACAGGGTCT GACTCTACGA CCCAAGCTGG AGGGCAGTGG      30500
TGCAATCTCA GCTCACTGCA AACTCTGCCT CCCAGGCTCA AGCCATCCTC
CCGCCTCTGT CTCCTGAGTA GCTGGAACTA CAGGCACACG CCACCACACC      30600
CAGCTAGTTT TTGTATTTTT TGTAGAGATG GGTTTTCACC ATGTTTCCCA
AGCTGGTCTC AAACTCCTCA AAAGAGTTAC CAAGCAGGAC TCTGCAACCA      30700
ATAATCCTTG TGTGAAGAGG ATATTTGCTC TTTTCCCTGT TTTTCTTTCT
TGGTACAGAT GTGTGACCTC TTTTTGAAAG GTGATAGTGA CTTTGGTGTA      30800
TTTTATTTGG TGGTAATGGT CATAGCCCCA TTAATCACAT TTCTTCCCAT
GAGAAAGAAA AACCACTACA TGGTCATGCT AAGGATTTCA GTCCCTGGGG      30900
TGAGGATGGT CTTGAATATC TCCTACATTC ATAACTCCTC CACACATCTC
AGTAGGTCAC TGAGCACATC AATGGACATG CCAGTTATTA AAATACTTCA      31000
CGAATACTAT GATCATTTAC CAGTATGAGT TATTCTCTGG AGCTTCTAAT
ACTTCAATAG TACTGCATGG ACTCAGTTGA GAGTTAATTC AAAATCTCAG      31100
      G
ATTATCCAAT TCTGTTTCTT TCCTTCCAGG CACCACCTAC CTATGATGCC
     [exon 11: 31130..
GTGGTACAGA TGGAGTACCT TGACATGGTG GTGAATGAAA CACTCAGATT      31200
ATTCCCAGTT GCTATTAGAC TTGAGAGGAC TTGCAAGAAA GATGTTGAAA
TCAATGGGGT ATTCATTCCC AAAGGGTCAA TGGTGGTGAT TCCAACTTAT      31300
GCTCTTCACC ATGACCCAAA GTACTGGACA GAGCCTGAGG AGTTCCGCCC
TGAAAGGTAC AAGTCTCCAG GGAAATGGAG CTCACCCTGA CCCAGGCTGG      31400
              ..31358]
TTCAAGCATA TTCTGCCTCT CTTAATCTAC ATGACAATCG TGTGGTTGTA
CAATCATTTG CTTGTAAGTC TTTTTATCAC AAAAAAGTGA TAATTATCAA      31500
ACTTTACAAA CCACAGACTA GAAAAAACGA AACTACATCC ATCCACAGTC
CCAGCACAAG ACAAAGATAA TCAATTATGT CCCTGTGGGC ATTTTTCTAC      31600
GCCTATATAG ATTTTTAAAA ATTAGAATGG TATCACTTTT TATTTGGTTT
GAATTGCTGC TTACTTGATT TAACAGGAAA CTATCCACTG ACCTATATTA      31700
CTATAAATAT ACATATATAT GTATATATAT AAATATATAT ATATGTATAT
ATTGCATATG CCATAAACCA TTTAACCATG ATGTTATTTC AGGTGTATAG      31800
GCTTTTTATT CCTTTCTGTT TTTTCTATGC TGTGCCCTTT AGCTCTCTGA
ATTTAACAGA AACTTTAAAA CATGCTTCCA CATTCCATTT GCTTTCAACG      31900
TTACTTGCTA TTTCCTCTGT AGTAATTATA AGAGTGCAGG CTGAGGTCCT
GAGAAGTCCT CATCCCTAAT GGTTTAAGCC ACTTCACTGA AGACACAAGA      32000
CAGCACAGGT CCTCCTGGTC CTATCTGTGG CTGCAGTCCT GTGCCAGCTC
CCTTATACTC TCAGTAGACA TCTCACACAC TCCTCCTTGG AGGTGTCTTG      32100
AGCATGCTCT TCTGGGAATT CAGGGACAAG GTCAGGCCTT AGGCACAGTT
CGCACTCTGG ATATAGTTGG TGTTTTCCCA TTACTGTATT ATTAAGCAAA      32200
```

FIGURE 1M

```
ATTTAGAATG AAATTTTTAG GGTACTGGCT GGTGATTCAG GATGCTTGGG
ATCTAGACTT TCATTAGCCC CTACCTGCAA GTTTGCTGAT GGGAGGAACC      32300
TTGTCTTGTT GGTCATGGTG TCCCTAGTGC TAGCATGGAG TCTGCACATA
ATACTTGTTC ACAGAGTAAG TCAGAGCTGA CCAAGTTCTC TGTTTTCTGG      32400
AGTAGAGGAC TTCTATGTTT CCTGCAAGCT CAGCACTTCC ACCTCCTGTG
GCTGCACTAA TACGAAATCA GAGACCACTC GCTGTACTTC ACTTTGAATC      32500
ACTCAGTCAC CAAAAAGATA GTGCTTGCCA TGTGTCAGGA ACTTGGCTAG
GCAGGGAGAA ATTCATATGA TTTATATAAA TCCATAAATC CATATGATTT      32600
ACATAAATCC ATAAATTCAT GTGATATATA CGTATATGTG TGTGTATATA
TATATTAGAG AATGTTTGAC ATATACACAA GTACATGTTA CCGACACCAG      32700
CCTATAGAAT AGTTTTCGTG CATCTCCATA TATCTATCAC TGGTTCCAAC
AGCCATCAAT CCATGTTAGC TGCCCCATCC AAATGCCACC ATCACCCTCC      32800
TCCTGACTAT CATGTTATTT TGAAGCAATA GCCTGTAAAT ATTTCAGAAT
GCTCTCCAAA ATATAAAGAC TCCTGTAAAA ACATATGACA ACAATGCCAT      32900
TATTACTTTC TTTGAATCAA CATTTTTTCC TTAATATAAT CAAATATTTA
GAAATCAAAT TTGAATAAAA CATGGGTCAA TCTTCAAAGA ATTTATAGCT      33000
TAATGGAACA GATCAAGGAA AGCAGGGATG ACACTACAGT AGGGTAGCAT
CATATGCCCA TGTAACTTAT GTGACTTAAA CTATCCTGTA AGGGTGTGGG      33100
GGAGAAAGAG AGGAAGAGAT GGAGAGAAGA AAAAGGAAGA GAAGGAGGAG
GAGAAGGAGG CAGAGGAGAA GGTGGACGGG GAAGGTAGAG AGGAGGAGGA      33200
GGGGAATTAG AAAAAAAGAG ATGACAGGAG AAGGAAAGGG AAAAATAACA
ACTTGAAATA GCACAAGACG TTTTCTCCTT CTCCTTTCTC AATGAGCATG      33300
TGACCAACAC AAGTGTGAGT TGAGGCAGGA ATCCACTTTT CCATCCATCA
GTCTTATCAT TTATGTGCCT TTTATAGTGT GAACACATCA CCACCCTGAA      33400
TATAATTTTA GTGTTTAGAG ATAAATATTA TTTGCAACAA TATTCATCTC
ATCTCAAGAA ACGCTCCTAT AGGGTATGGA GAATTTAAAG GACCTGTAGG      33500
TTATGATGAT TATAACGAAA TAACCAAAGC AGGATTTCAA TGACCAGCCC
ACAAAAGTAT CCTGTGTACT ACTGGTTGGG AGGTGGAGGG GGGTTGTTCT      33600
TAAGTAAGAA CCCCTAACAT GTAACTCTGT GGTTTTTATG TTTCATTAAC
                                          A
TATTTAATCT ACCAATATGG AACTAGGTTC AGTAAGAAGA AGGACAGCAT      33700
     [exon 12: 33679..
AGATCCTTAC ATATACACAC CCTTTGGAAC TGGACCCAGA AACTGCATTG
GCATGAGGTT TGCTCTCATG AACATGAAAC TTGCTCTAAT CAGAGTCCTT      33800
CAGAACTTCT CCTTCAAACC TTGTAAAGAA ACACAGGTCA GTACACTTTC
                ..33836]
TGTATGTTTT ATTAAGAATT TTTTTAACTG AAGGGTATAT ATTTTTTAAA      33900
AGAATATGCA TGTTTATCTT TTAATAATTC ATTCTATGGG CCAAAGAACC
TACTTGGATC CATCTTTGAT CATTAAGGAT GCTTCAGTTC TGGACTTCAA      34000
AACCTGTAGC ATTAAGAACA TCATGTAAAG TCCACACAGA TTAGCATGAC
ATGATTATGT GTAGTCTCTT TGAACCTGAG TAAGTTTAAA TTCAGTTTCA      34100
AGTCAATTGG AAAGAAGTGT TTTGCACAAT CATGAAGTGC AATGATTACC
TGGCTGTGAC TTAAATGGTG TTCTCCATCA CCAGAACCTG CAGAAGCTCT      34200
CTCATGACAG TGGTTCTCAA CCACTAGCTG TATATTGGAA TCACCAGGGA
GCTTCAAAAA TTCATGATGC CTGTGACATC TCAGAAATTC TAAACTAATT      34300
AACCCAGAGC GTGACTAGGT TCTGTCATGC TGTCGGGTGA ACCCCTGATT
AGTTCTCACG TGAAGCCAAG GTGGAGAATG ACTAATTTCA GGCATTTCTG      34400
GTGGATATGA AGGACTACCA TAGAGCAGGG CTATCCTTAC TCCTTGACCT
TATGTTCCAG GTGATACATT TAAAGAAAGA TTTAGAATCT TTTCTCTGAA      34500
GAAGTTAAAG AACAGATGTC ATTGATTCAT ATTAAGCAAT AGCCTATAAG
TCTTATTTCC AGGACCGGTG TATTTAATAT GCAACTCTAC CCCTTAAGTA      34600
CACTTTGTGC TTGGGAGAGG AGGAGGATGG AGATGGTTGC CATCTTATCT
ATGGCTTCAG GGCAGCTGTG TAGCTTTCCT ATGTGTGTAT TCAGGCAGGG      34700
```

FIGURE 1N

```
GGCTCAGCCC TGAGAGAAAG TGGGCCTCTG GCACACCTGG GACAGGGAAG
ATATTCCCTG GCAAGCTCTC AGGCATCTCA GGCTGGCACT TCTTTGTATC      34800
CATGGCAATT TGCTTTCCCC TCACTGAACT GAGATCAGAA TGTTACTCTG
TTGGTGGCTC CCCCAACAGT GAAGGGGTGA CTCAGTGACA ATAGTGCTAG      34900
AAGTATGAGT CAAAACACTG TACAACTTGA GAAATTCCCC GTTTGCACTA
CGCTTGGAAG CCAAGAGGAG ATGTTAAAAA GAAAAGAATA ATTCTTTCTG      35000
AAGACATTTC CCATCATTGC ACTTGATGGG TTCAACTGGG AAGGGTTACT
AGACTCTGGA AGTTGAAAAC TGCCCACATA ATTAAACTGT ACAACAGCTA      35100
CTCAGGATTA CCTTGCAAGT TTTAACCTAT AAAAATTTAA CTTTATATAG
CACTTCCAAA ATAGTTTGCC ATAATACCTA CTAATCTGGA TTTAATTTTT      35200
AAAACTCATC CTTTTAACTT AAGATTTAAA TAAAAAAAAA AAAACACGAG
TCCACAAGAA TTTGTCTCAG GCCTGGCACA GAGTCAGTGC TCCATAAATA      35300
TTTTGTTAAA CGATGGATGG TGAGTGCTTT TACTATCCAG TATTTACCCA
GCTTATAGAT TAAGTATGAA GAGTTCAAGA TACATGGTGT TAAGAGTCGT      35400
TTTTATATGC TTGCAAAGCA TTTTTGTCAT ATTTTTTCTA CTTTGCTTCC
ATCTTTTCTT CTTTCACTTC ATTTATTAAT TCTCCATATG CTTGTTTAAC      35500
TATTGTAGAT CCCCTTGAAA TTAGACACGC AAGGACTTCT TCAACCAGAA
    C
     [exon 13: 35509..
AAACCCATTG TTCTAAAGGT GGATTCAAGA GATGGAACCC TAAGTGGAGA      35600
ATGAGTTATT CTAAGGATTT CTACTTTGGT CTTCAAGAAA GCTGTGCCCC
                C
               ..35604]
AGAACACCAG AGATTTCAAC TTAGTCAATA AAACCTTGAA ATAAAGATGG      35700
GCTTAATCTA ATGTACTGCA TGAGTAGTTG GTGATTTTGT ACATTCATTG
AGCTCTCCCA GAGTCTGTGT AGAGTGTTGT GCATTATGTA GTATAAAGGA      35800
GGTGACCAGG TAAGTGACAG ATAGGTAGAC TCAGCTTCTC TGCTTCTCAT
AGGACTACCT CTACCCACCT CTAGTTAGCA TTATCAACTC CTCCTGAGCT      35900
CTCATCAGAG AATAAATATT TCTCAACAAT TTGATCCATA ACTTTTAAGA
AAAATAAGAA TTATCATGAT GACTCTAATA GTGACATTTA TATCACGTTT      36000
TATTTGTAAT ATTCTATAAG TTTTATATTA AGCGAAGTGA TAAAATCCCC
TTTACAAAAA TATTATCTGA TGCCATCCTG CACACTAAAG AGAAATCTAT      36100
AGAACTGAAT GACTGAAAAC CAGCAAATAA ACATTTTTTA TCATTGTAAT
CACTGTTGGT GTGGGGCCTT TGTCAGAATT CCAATTTGAT TATTAACATA      36200
GGTGAGAGTT AATCTGCTGT GACTTTGCCC ATTGTTTGGA GAAAATATTC
ATAGTTTCAT TCTGCCTTCT TTGAAGAACA TATTTTTTGT AACACTCAAC      36300
GAAGCACTTA TCATATTATT AGTTATGATT TATTATTTTT ACCACATCTC
CCCTGACATT TCTGGAACAC AGGAAACATG TTTTCTTATA CGTCTTGCAT      36400
TCCATCTTCA CCTCCCAATT GTCTTAATGC AATGAACACT GAATAAAAAA
TTGTCAATTC GTCAGTTGAT TGGGCAGCAT GTCTAAAAGC ACTATTCATT      36500
TTCCTTTTTT ATTCTTTCAT TTTCCCTCCT TTTCTGAATA CTAAAGCCAT
TAGGTGGGTT GCAGCCATGT GGTAGCCACA CATTAAGGTG GACAAGAGAG      36600
TCATGGTGGC TCCAAGTCAG ATTCCAAGTG TGCTGGGGAA GGCATCCACA
TGGAGGGGCA GCCTGACCTG GAAGCGGGAG CCCAAGCAAT CAGAGAAGGG      36700
GTCCACACAG AGGTGTGGCC TTCAAGAGCA GCCAGAGCCT AAATAGGGCC
TGGAGAACCC ACGTGAGGTG AGGAGGGTAT CCCTGAGTGG GAAGGGATGG      36800
GTGAGAGTTG GCTACATAGA AGGGATTGAT CACATAAGTA AATAAAGTAT
ACTGGAAGCT AGGTGTGTCA CTTTTGCAGA AAAGAGTCAT AGATTCAGAA      36900
AG                                                          36902
```

FIGURE 10

POLYMORPHISMS IN THE CODING SEQUENCE OF CYP3A5

```
ATGGACCTCA TCCCAAATTT GGCGGTGGAA ACCTGGCTTC TCCTGGCTGT
CAGCCTGGTG CTCCTCTATC TATATGGGAC CCGTACACAT GGACTTTTTA      100
                                      T
AGAGACTGGG AATTCCAGGG CCCACACCTC TGCCTTTGTT GGGAAATGTT
TTGTCCTATC GTCAGGGTCT CTGGAAATTT GACACAGAGT GCTATAAAAA      200
GTATGGAAAA ATGTGGGGAA CGTATGAAGG TCAACTCCCT GTGCTGGCCA
TCACAGATCC CGACGTGATC AGAACAGTGC TAGTGAAAGA ATGTTATTCT      300
                                                  A
GTCTTCACAA ATCGAAGGTC TTTAGGCCCA GTGGGATTTA TGAAAAGTGC
CATCTCTTTA GCTGAGGATG AAGAATGGAA GAGAATACGG TCATTGCTGT      400
CTCCAACCTT CACCAGCGGA AAACTCAAGG AGATGTTCCC CATCATTGCC
CAGTATGGAG ATGTATTGGT GAGAAACTTG AGGCGGGAAG CAGAGAAAGG      500
CAAGCCTGTC ACCTTGAAAG ACATCTTTGG GGCCTACAGC ATGGATGTGA
TTACTGGCAC ATCATTTGGA GTGAACATCG ACTCTCTCAA CAATCCACAA      600
GACCCCTTTG TGGAGAGCAC TAAGAAGTTC CTAAAATTTG GTTTCTTAGA
                          A
TCCATTATTT CTCTCAATAA TACTCTTTCC ATTCCTTACC CCAGTTTTTG      700
   G
AAGCATTAAA TGTCTCTCTG TTTCCAAAAG ATACCATAAA TTTTTTAAGT
AAATCTGTAA ACAGAATGAA GAAAAGTCGC CTCAACGACA AACAAAAGCA      800
CCGACTAGAT TTCCTTCAGC TGATGATTGA CTCCCAGAAT TCGAAAGAAA
CTGAGTCCCA CAAAGCTCTG TCTGATCTGG AGCTCGCAGC CCAGTCAATA      900
ATCTTCATTT TTGCTGGCTA TGAAACCACC AGCAGTGTTC TTTCCTTCAC
TTTATATGAA CTGGCCACTC ACCCTGATGT CCAGCAGAAA CTGCAAAAGG     1000
AGATTGATGC AGTTTTGCCC AATAAGGCAC CACCTACCTA TGATGCCGTG
GTACAGATGG AGTACCTTGA CATGGTGGTG AATGAAACAC TCAGATTATT     1100
CCCAGTTGCT ATTAGACTTG AGAGGACTTG CAAGAAAGAT GTTGAAATCA
ATGGGGTATT CATTCCCAAA GGGTCAATGG TGGTGATTCC AACTTATGCT     1200
CTTCACCATG ACCCAAAGTA CTGGACAGAG CCTGAGGAGT TCCGCCCTGA
AAGGTTCAGT AAGAAGAAGG ACAGCATAGA TCCTTACATA TACACACCCT     1300
TTGGAACTGG ACCCAGAAAC TGCATTGGCA TGAGGTTTGC TCTCATGAAC
ATGAAACTTG CTCTAATCAG AGTCCTTCAG AACTTCTCCT TCAAACCTTG     1400
TAAAGAAACA CAGATCCCCT TGAAATTAGA CACGCAAGGA CTTCTTCAAC
CAGAAAAACC CATTGTTCTA AAGGTGGATT CAAGAGATGG AACCCTAAGT     1500
GGAGAATGA                                                  1509
```

FIGURE 2

ISOFORMS OF THE CYP3A5 PROTEIN

```
MDLIPNLAVE TWLLLAVSLV LLYLYGTRTH GLFKRLGIPG PTPLPLLGNV
                              Y
LSYRQGLWKF DTECYKKYGK MWGTYEGQLP VLAITDPDVI RTVLVKECYS      100
                                                    Y
VFTNRRSLGP VGFMKSAISL AEDEEWKRIR SLLSPTFTSG KLKEMFPIIA
QYGDVLVRNL RREAEKGKPV TLKDIFGAYS MDVITGTSFG VNIDSLNNPQ      200
DPFVESTKKF LKFGFLDPLF LSIILFPFLT PVFEALNVSL FPKDTINFLS
KSVNRMKKSR LNDKQKHRLD FLQLMIDSQN SKETESHKAL SDLELAAQSI      300
IFIFAGYETT SSVLSFTLYE LATHPDVQQK LQKEIDAVLP NKAPPTYDAV
VQMEYLDMVV NETLRLFPVA IRLERTCKKD VEINGVFIPK GSMVVIPTYA      400
LHHDPKYWTE PEEFRPERFS KKKDSIDPYI YTPFGTGPRN CIGMRFALMN
MKLALIRVLQ NFSFKPCKET QIPLKLDTQG LLQPEKPIVL KVDSRDGTLS      500
GE                                                          502
```

FIGURE 3

CYP3A5_1385.ST25.txt
SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Anastasio, Alison E
Han, Jin-Hua
Kliem, Stefanie E
Rounds, Eileen

<120> HAPLOTYPES OF THE CYP3A5 GENE

<130> CYP3A5_MWH-1385PCT

<140> TBA
<141> 2001-12-07

<150> 60/288,470
<151> 2001-05-03

<150> 60/254,367
<151> 2000-12-08

<160> 109

<170> PatentIn version 3.1

<210> 1
<211> 36902
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (3633)..(3633)
<223> PS1: polymorphic base A or G

<220>
<221> allele
<222> (3747)..(3747)
<223> PS2: polymorphic base C or G

<220>
<221> allele
<222> (3927)..(3927)
<223> PS3: polymorphic base G or A

<220>
<221> allele
<222> (3939)..(3939)
<223> PS4: polymorphic base C or T

<220>
<221> allele
<222> (3998)..(3998)
<223> PS5: polymorphic base A or C

<220>
<221> allele
<222> (7657)..(7657)

<223> PS6: polymorphic base T or C

<220>
<221> allele
<222> (7717)..(7717)
<223> PS7: polymorphic base C or T

<220>
<221> allele
<222> (7830)..(7830)
<223> PS8: polymorphic base G or A

<220>
<221> allele
<222> (9523)..(9523)
<223> PS9: polymorphic base T or A

<220>
<221> allele
<222> (11189)..(11189)
<223> PS10: polymorphic base C or A

<220>
<221> allele
<222> (11214)..(11214)
<223> PS11: polymorphic base C or T

<220>
<221> allele
<222> (11310)..(11310)
<223> PS12: polymorphic base C or A

<220>
<221> allele
<222> (16830)..(16830)
<223> PS13: polymorphic base C or T

<220>
<221> allele
<222> (17383)..(17383)
<223> PS14: polymorphic base G or A

<220>
<221> allele
<222> (18697)..(18697)
<223> PS15: polymorphic base G or A

<220>
<221> allele
<222> (18727)..(18727)
<223> PS16: polymorphic base A or G

<220>
<221> allele
<222> (18787)..(18787)
<223> PS17: polymorphic base C or T

<220>
<221> allele
<222> (19755)..(19755)
<223> PS18: polymorphic base C or T

<220>
<221> allele
<222> (19806)..(19806)
<223> PS19: polymorphic base T or C

<220>
<221> allele
<222> (20065)..(20065)
<223> PS20: polymorphic base A or C

<220>
<221> allele
<222> (21170)..(21170)
<223> PS21: polymorphic base G or T

<220>
<221> allele
<222> (31057)..(31057)
<223> PS22: polymorphic base A or G

<220>
<221> allele
<222> (33640)..(33640)
<223> PS23: polymorphic base G or A

<220>
<221> allele
<222> (35506)..(35506)
<223> PS24: polymorphic base T or C

<220>
<221> allele
<222> (35618)..(35618)
<223> PS25: polymorphic base T or C

<400> 1
ttactttccc ttcctgagta acttatccta aagtcattag gtgggtggca gccagatggt 60
ggccacacat taaggtagaa aagagagtgt catgatggtt ccaagtcaga gacctagtag 120
ggtgaggatc aagtaggtgt tcacgtggag aaacagcccg gcctgtgtgt gggagtccaa 180
gcaagcagag aaaatgtcga cacagagggg tggcctgaaa aagcagccag agcctaaaca 240
gggcatggag aacatattta gggcatgagg tgaggagggc atccatgagt gggaagggat 300
gggtgaggtt tcactacata aaggggattg atgaaataag taaataaagt atactggaag 360
ccaggtgtgt cacttttgca gaaaagagtc atggattcag aaagggagaa aactagcagg 420
aatcctatga aattagatta aaatggatgt atccatgtat attcataccc ttctagatag 480

ataaatggtt agataggtga taaaaagata acaagaggac aagataatta gatagacata 540
aatgtatgta tgtgtttgtg tgtgtgtaca aaaaaacata tactccctac ttctctccac 600
tgatagggct aggtaacaat ggcatttcaa tagcaatgag cacacttagt ggccagatct 660
tggcttatta ataccatttt ccactgaaag gaaccagagc tttttagaga aatggctgat 720
tccagggcca ggattaagaa tgttcaagat aagcctagga tacattttgt gccaggaagc 780
aagaagatgt tcaaatgatt tccaagtaat gtttggaaat gatatttgaa aatgatttcc 840
aaatgatatt tccaaatgat ttccaaatga tatatggaaa cacttaaaga ctccactaaa 900
gaactattag atctgataaa caaattcagt aatgttgctg gatacaaaat caacatacaa 960
aaaccagtag catttctgca tgccaacagt gaacaatctg gcaaaaataa aaaatgtaat 1020
cccatttaca ataaccccaa ataaaactaa atacctggga attaacttaa gagaaagatg 1080
tctacaatta atattgtaaa acactgatga aggaaattga agaagacaca aaaaagaagg 1140
atattccatg tttatatatt gtaagcatta atattgttaa aaatgtccat actacccaaa 1200
gcaatgcaca gattcaatgc agtctctcaa aataccaatg gcattcttca aagaaataga 1260
aaaaaaaaaa ccctaaaatt tgtatggaac cacaaaagac ccagaatagc gaaagctacc 1320
ttcagcaaaa agaacaaaac tggaggaatc atattacctg acttcaaatt atactacaga 1380
ggtataataa ccaaaacagt atggtacttg tataaaaaca gacacagacc aatgaaatag 1440
aatagagaac ccagaaacaa ttccacacac ctacggtgaa ctcattttca acaatgttgt 1500
caagaacata cactggggga aaagacagtc tcttctggtg ctgggaaagc tggattttaa 1560
catgcagaat aatgaaacta gaaccctgta tctcaccaga cacaaaaatc aaatcaaggt 1620
ggacgaaaga ctgaaacctg gctgagtgcc gtggctcatg cctgtaatcc cagcattttg 1680
agaggccgag gcgggtgtat cacttgaggt caggagttca agaccagcct ggccaacatg 1740
gtgaaaccac atgtctacca aaaaatacaa gagttagctg gacatgctgg tgcgtgcctg 1800
tagtcccagc tacacagaag gctgaggtgg cagaatcact tggacccagg aggcggaggt 1860
ggcagtgagc tgagatcatg acaatgcacc ccagcctggg caagagagtg agactctgtc 1920
agaaaaacaa aaaacaaaaa aacaaaaaac aaaactgaaa tctgagacct caaacgatga 1980
aactgctaca agaaaacatt gtggaaactc ttcaggatat tggtctgggc aaaactttct 2040
gaagaactac cccacaagca caggcaacca aagcaaaaat ggacaaatgg atcagatcaa 2100
gttaaaaagc ttctgtacca caaagaaagc aatcaacaaa gtgaagacac aaaccacaga 2160
atgggagaaa atattttcaa agtcacactc tgacaacaga ttaatagcca gaatacatga 2220
agcgctcaaa caactctgta aggaaaaatc taataatcca atcaaaaaat gggcaaaatt 2280
tgaatagaca tttttcaaaa gaagacatac aaatgccaca taggcatatg ataaggtgct 2340
caacatcact ggtcattaga gaaatgcaaa tcaaaaccac aatgagatat catcttaccc 2400
cagctaaaat ggtttttatc caaaagacag gcaacaacaa atgccagcga gaatgtggag 2460
aaaagggaac ccttgtacac tgttggtgta aattagtgca accactatag agaacaattt 2520
ggaggttcct caaaacatta aaattaacat taaatagagc taccacaata tccagaaatc 2580
cccatgctgg gtatatacct ggaagaaagg aaatcatata ttgaagagat aacatcactc 2640
caatattcac aatagccact attcacaaat gccaagattt ggaagcaacc taagtgtcca 2700
tcaacagatg aatggataaa gaaagtactc caattataca caatggagca caattcagcc 2760
atgaaaaaag catgagatcc tgttatctgt aataatatgg atggaactgg aggtcatcat 2820
gttaagtgaa ataagccagg cacagaaaca cagatattgc aagttctcac atacttgtgg 2880
gatctacaaa tcaaaacaac tgagctaatg tctgggcctt agtcagtgtt gtacccaagt 2940
actgggagca cagcttttaa aatacatcat gaatgcttta atacaggaat gaatagatga 3000
gaggcacaaa ctggttgggt gttcttctga tacacagtat cttccttgac agattcagta 3060
caactctcaa caggtaagtc tcttcatgtt atgttacctt atgaggaatt aagtggcaga 3120
acatgatttc tattattttc ctttgcagaa caagaccaac tttattagtt gggacacagt 3180
gtggctgcat ttgagtccca agcaaccatt agtctattgc tatcaccaca gagtcagagg 3240
ggatgagacg cccagcaatc tcacccaaga caactccacc aacattcctg gttacccacc 3300
atgtgtacag taccctgcta ggaaccaggg tcatgaaagt aaataatacc agactgtgcc 3360
cttgaggagc tcacctctgc taagggaaac aggcatagaa acttacaatg gtggtagaga 3420
gaaaagagga caataggact gtgtgagggg gataggaggc acccagagga ggaaatggtt 3480
acatttgtgt gaggaggttg gtaaggaaaa attttagcag aaggggtctg tctggctggg 3540
cttggaagga tacgtaggag tcatctagag ggcacaggta cactccaggc agagggaatt 3600
tcgtgggtaa agatgtgtag gtgtggcttg tgrggatgga tttcaattat tctagaatga 3660
aggcagccat ggaggggcag gtgagaggag ggttaataga tttcatgcca atggctccac 3720
ttgagtttct gataagaacc cagaacsctt ggactccccg ataacactga ttaagctttt 3780
catgattcct catagaacat gaactcaaaa gaggtcagca aaggggtgtg tgcgattctt 3840
tgctattggc tgcagctata gccctgcctc cttctccagc acataaatct ttcagcagct 3900
tggctgaaga ctgctgtgca gggcagrgaa gctccaggya aacagcccag caaacagcag 3960
cactcagcta aaaggaagac tcacagaaca cagttgamga aggaaagtgg cgatggacct 4020
catcccaaat ttggcggtgg aaacctggct tctcctggct gtcagcctgg tgctcctcta 4080
tctgtgagta actgtccaaa ctcctctctt tgtttccttg gacttggggt gctaatcggg 4140
ccccttttcc cttatctgtt ttgaagatca aaagagatgt tcaaggagaa gtagctgaag 4200
tgttggacgc tacaaacgca tagaagttat tattatctta tgcagatcta tgaatgaata 4260

aataagcatt tctcccatcc accttctaat tttggtgact aggagggttt agggacagca 4320
tttggtagtg ggaatgattt gattagctta gatctgacga agactaatca atgaaaacat 4380
ggcagcggca gattacaaac tgctgatcat gatggacagt gtgatcctca tccccttccc 4440
aggctctggg gattctgggt acaggaagga gtggcttgca tttttgtctc attaattcgc 4500
tttctgggtt ctgtgtctgc tggaagggat gtgtagctgt attgcccctg tagacctggt 4560
tcctgctccc ccgccttcca acccaggata tcatttacat aacgcaccag gggacaccaa 4620
gacttcatgg gaagctgtcc cctggctctt ccctctttcc tgtgccatgc ccctgaaaat 4680
cccctccctc ctatgagtca ctcctccacc ctgtcataca caggatggtt tatcttgcaa 4740
tgattaacct ctagagcaaa ggagacctgg aggaagtttc gaggatttat tctttgcttt 4800
aatctttttc ctcccgtctc tgggaggcta ggattaatat agagctttgt ttctcaccta 4860
atgggaatct actagcagcc tgaaaaggca ggagccatga aagccaattt ggattttaca 4920
tatttttccc ctttatgtta cagtacagga gggcaaaccc tctcactggt gggattcctg 4980
gcatcctaga gcaggtggag agaagagtta ctttccactg tgggtagtgg aggctccacc 5040
tgtcccatta acttctacct caatttgact tttattaaga gcagggaacc acaatgacat 5100
gaaaatagac actataaacc tcattttaat tctttcacag aaagcttagg aattcagtga 5160
gttgtggcaa catggtttcc attgtctaac atttttaaat gaattgatat ggtttaaatt 5220
cattcatttt taaaccagaa ttttttggag atagactatt tccagcatgt tccttctgga 5280
tggtaaaaca gggctgttag ttcagtattt gtgacaataa gtgtgtgtaa aataatgtca 5340
cctttcctga atgtcaggaa tatgagtcta atgcacaaat gtatacctct aagacaagac 5400
tgcacgtctt ttcaaatata cctgtccggc catttatttt aataactcct tttcgaatat 5460
acctgcttag cagattgtct taaactctca ggacagggga gtaagcaaga ctgtgagcca 5520
gtgacgatag caaaggcttc caggtaggat ccatatgaag tgagaaaata ttcctcagct 5580
ctcagggtag aactccaaag agatattcat gggtcctggc cccaccgtgg aggtcactca 5640
aagggcaaac aggttggcat ctcatctgct tcaagcctgg acacaggggc accatctgtg 5700
tcactctgtg tgtggtctgc catgttgtgg gccggtcact acagactcgg gcagccaggc 5760
agacaatgcc ttagccttag acaatgctgg tgcagcccag gagtcagaaa atgcagtgta 5820
gaccaggccc tccttaggcc aacacaatta catgcaatag atgactggct tttctgttag 5880
tctcttcact ggacccaaag gctgcattac tctaccagag gggagctgga aagaaactaa 5940
agagttcgcc cagcacagca tctgccttga catggtacca tgtgaatcta gacactcacc 6000
aagatctttc cttgggggcc aatgctgctg acacattaac tcaatagctt gtcctcacct 6060
gagaggtcag gtaatgtgtt taaagttcag gagcagagat tagtgtcatt gatttgacat 6120
ggctgtgaca acaaaggagg gaactgaagt gggaataccc aaggccaccc tggctttggc 6180
aggtggtgca cgcacttcca ctaactgttc tggggcaggg aaccaaatgt atgactgggc 6240
ctgctcatgc tgcccctgct gagtcctcca aaccctgccc ttcatgtaat ttctcagttt 6300
tattttatca cattttataa gtcactggat gtttacaaaa tgtttggaac ctatactgcc 6360
ttgaaggcta acctctaaag aggagtaaac aaggtcttaa tacaactctc cgggacgttt 6420
tatcattact tatcttatat gccatactgc accatttgct atcaacagga aagtacctgg 6480
actttggaag gtccctctgt gtcttttagc tgaaagtaca tatgaggcat gtggattctt 6540
ttatgcacat catctttttc agccacattt ttgtagtttg cctctctgga gccaactgtg 6600
tggggctagc agcttcacag ctgaatcagt gtctggcaac ctcttccttc agcctctctt 6660
cttcctccag ttttccatcc ctcagtcaca ccggaggggg aaggtctgca aggatccaga 6720
accatcagtt ggaggagttt gcacatgact catgaaagat gagttccagg caggcctgcc 6780
atagtgaaca ccaggcttaa tgggtttttc ctcagagata cttcacgtac agaggcagtg 6840
aactgactgc tttctggttg accaccttga aaaagatgag tgtgcctggc actgtgcttc 6900
tcaggtgagt atgacctgag aagtattagt tgctggttct tctgcacaca atcattcaag 6960
gacatatgga tcaaccatcc tcctcaacag ctcaaatcaa ccagatcatc tgaccacaga 7020
gactgaggtg tacctgaaag ctgcccacat ttctataagg ccaatagaag ccatgaacac 7080
agttgtcaat ctgtagaaat aaggactcca tgactcctcc aaggcctctc tgtgaatgaa 7140
cgtttaagaa gggctagatc ctaaaacagg gtcagagctt agagggaaga aaaagcataa 7200
acatttctga gcaaattgta agggcagtgt caccataggc tcccagtgac cctctgtgat 7260
tgagtgcata cagtgatgca aaatctcatc atcagtgcaa aagacaaaaa aaatcttact 7320
ctttctacct aggatgagag tccccaaatc agcgaagagt ccacttacta aacagacata 7380
aggaaatgaa gtgtcctgga agaattcctg cctgaacctc tcaggagcat ttgaggacat 7440
ttatcaagta ttcactccag gattgggact atgaagactt cagctgcttt cagctaatca 7500
ttgagacttt tcaggggtct cagaatagtc aggaaaggac ctgatgagtg aatgcaatta 7560
ctgatgttgg agttgctgtt attatttatc gtgtacatat tacctccctc tcttgaccat 7620
tccagttcct gagtaactca ccagccctct gatctayaaa gtcacaatcc ctgtgacctg 7680
atttctgttt cactttgtag atatgggacc cgtacayatg gacttttaa gagactggga 7740
attccagggc ccacacctct gcctttgttg ggaaatgttt tgtcctatcg tcaggtgagt 7800
tgcttgagct tcctcttttg cttcttatgr ttgcaaacat cagcttagtt ccatcagtaa 7860
aaatgcccct ccttgggagg gagttctgag gtttcacatt ttcagaaatg gtgggactgg 7920
gtgcagtgga tcatgcctgt aatctcagcc tctgtgaggc caagactggc aaattgcttg 7980
agcccaggag tttgagaaca gcctgggcaa cacagtgaga cacctgtctc tagaaagaaa 8040

aaattacctg tgcatgatat ggtagcccat gcctgtagtc ccagctactc tgaatgttaa 8100
ggtgggagga ttgtatgaac ccaggaagtc aaggctgtat tgagctgtga tcgcaccact 8160
gcactccagc ttggtcaaca gaacaagaca gaaaggaaga aagaaagaga gagagagaaa 8220
gaaagagaga ggaaggagag gggaggggag gggaggggag gggggaggag aggagaggag 8280
agaaaaggag aggagagagg agaggagagg aaaaggtgtg taggctccac ccaaagcatg 8340
gccaggttta cccctggagg gaaagtcaca agctcatgtc cagaaggcca gtagcagcaa 8400
gctgctctcc agcccagatt tcctatcctg tgtacctgga gcttgtttct cagattctaa 8460
ctctcacaac tgaagcctct gttgtctgat tactatctga gaattctaca caattttacc 8520
ctcgataaaa gcagtaattt cttcttcatc tttcccagat caactcttgt agtagatcaa 8580
catttctggg accttctttt gcatggttaa aacatcacag ctgaatctta gcaacaggaa 8640
ggtttgtttt tatgtttcag aagtgaaagc tcagagcacg cattgtaatt tgctgggtgt 8700
gatgtgtaga ggtggcattt ctccatcttt tctgtgttaa gctagaaaac tggaaaggaa 8760
gtctactttc tcattcactc actcactttc tcactcaaca acatgcctta gacttatcta 8820
aatctgcaag actaaaagag gttcctggtt tctttaactt tctaattctg ctagagttct 8880
agagagagca catgagataa atgaaaagga tactgatgga ggagattaaa aaattgtgca 8940
ttccctgcag acactcactt ttcctcacct cagtttcacc cctgcccttg caggtgatca 9000
ttcacggggt taggagactt tagagagaat aaaagaaaaa gcaaaaatac atcagaaaga 9060
caaggaatta cttactggtc atagacaagg gtgagtcctt cagtacttag agaaaattca 9120
agagtgactt taaattcccc acttcaaata tattctctgt tttcttgtct ttcccttaag 9180
acatctctga atagcttcct tcaactgcca gtgaaagata gcaggcctga tttcattgga 9240
cgcaactgtt ttcagcccca attagaggta gggtttattc tatttaaaat aataatcaac 9300
ttgtattttg tttcctctcc cagggtctct ggaaatttga cacagagtgc tataaaaagt 9360
atggaaaaat gtgggggtga gtattctgaa aacctccatt ggatagacct gctactgtga 9420
ggaggttacc ccactgcagg atagtctctg cccaggtctt catgggatga agctcttgtc 9480
aacctaaata caaacagaga gaggttctct gaaagaagag gawaattact tgggagtaga 9540
atattgcaat gggaatctgc ttgccgttat aaactatgtg caaattcagg gaggtaaaca 9600
agacaaagat gctccataga aaatatgaga agaatctcat aactgttttg agataattat 9660
tgttagctac aaagatcaat aacaagggtg atgccacacc aaggttggac aggcagttgc 9720
tggacaggtg tccttgcaga aatatttttg tgtaaagttg aaatagcctt tgtgcaaagt 9780
tgtggttttt gtagacactt ttgtaatagt tttgtttcca ggaacacaag cataagaatc 9840
ctctcttcat agccttcttg ggatttattt gtcagggtta aaaaacaatt agtgacatca 9900
ctttggttct gataaagttc acactcgcta ttgtaaaact tttcgaggct tgtcctacca 9960
aggatcccat gtgtcaccag gtatcgaggt cttcagtctg aactaggcta ggagcattgt 10020
ggttaccact tttctgcagg ttttggtggc ccagggactc ccagcatcgc cttctgtcca 10080
gtgtctgcct attcccctct tctttttttc ttccttaggt gccctttcat cacatgcatt 10140
gtctcagacc cttctaatat gtgctcataa atgcatggca tcatctcctt cccacattga 10200
ttcactttca attaaaagcc aaaactcctt catttagact gaatttaaca tgtgcttttg 10260
aaagaagggt tgagagataa tagagaaaca gattgggaaa ccacttatgc tccacttttt 10320
taaactttct ctgcaagtat ggaatttttt gttctgcttt gttgtttaaa tttaagccaa 10380
aacttcttaa tagaaggata tacaaatatt tattggttta taccattgca cttactttga 10440
agaagagatg ctgaatatta ttaaaccatt gtgttccctg gtgggctgat ggactgtgat 10500
tttataaggt ggtctcagcc aattgcagca gctgttccct gtcagagggg ctagaggttt 10560
ggtgagagca gtggatgagg tgcagtggtg tgtttgttca ctagaagcaa gtgggagaaa 10620
gctttgcctc tttgtacttc ttcatcttct cccctcaagt cctcagaatc cacagcgctg 10680
actgtggagt gctgtggagc tggcatggcc catacaggca acatgactta gtagacagat 10740
gacacagctc tagatgtcca tgggccccac accaactgcc cttgcagcat ttagtccttg 10800
tgagcacttg atgatttacc tgccttcaat ttttcactga cctaatattc tttttgataa 10860
tgaagtattt taaacatata aaacattatg gagagtggca taggagatac ccacgtatgt 10920
accacccagc ttaacgaatg ctctactgtc atttctaacc ataatctctt taaagagctc 10980
ttttgtcttt cagtatctct tccctgtttg gaccacatta cccttcatca tatgaagcct 11040
tgggtggctc ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag 11100
gttgctgtgt gtcgtacaac taggggtatg gattacataa cataatgatc aaagtctggc 11160
ttcctgggtg tggctccagc tgcagaatmg ggctagtgaa gtttaatcag ctcygttgtc 11220
cccacacaga acgtatgaag gtcaactccc tgtgctggcc atcacagatc ccgacgtgat 11280
cagaacagtg ctagtgaaag aatgttattm tgtcttcaca aatcgaaggg taagcatcca 11340
ttttttgaaa tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaacat 11400
atattcacag aaggttacct aaaaaatgta caggaaggtt ccatgtactc ttcatcctgt 11460
cccgcccagt ggtaacatct tgcaatcttg tatattgcaa tatatatcta gtatattcat 11520
attatcaggt tggcacaaaa gttaaaatgg caaactacag gctgggcata atggctcatg 11580
cctgtaatcc cagcactttg ggaggccgag gcaggtggat cacgaggtca ggagttcgag 11640
atcagcctga ccaacatggt gaaaccccat ctctactaaa aatacaaaaa ttagctgcgt 11700
gtggtggcat gcgcctgtag tcccagctac tcagtagtct gagacaggag aatcgcttga 11760
acctgggagg cggaggttgc agtgagccga gatcacgcca ttatactcca gtctgggcaa 11820

cccaatgaga ctccatctca aacaacaaca acaacaacaa caacaaaaac cggcaaactg 11880
caataacttt tgcaccaacc taatactata gtacaggaaa ttgactttga tatagtttac 11940
agagcttttc agatttcacc agttttacat gcccttgttt gtgtgtgttt atgtgtgtgg 12000
gtagttctaa gcaatttttc acattcgtag atttgtgcaa cgaccagcac catcaagatg 12060
cagacccatt ccgtcaccat gtggctccct cctgctgtcc tacagtcaca acatggagtt 12120
tgtctttttc tctgacaggt tctatatcag agcaaacttt tatttatttg aggaggccaa 12180
tgtattaata tttccttta tggattgttc ttttggtgtt aagtctgaaa atcctttgct 12240
tagccctcct tcctacattg ctttttctaa gagttatata gtttaacact ttacaaaatg 12300
taactctatt acccattttg tgttaatatt tgcataagtt atgagattta gatcaaggtt 12360
cattttctgt ggactatggc tgtccaaatg ttccaacacc attttggaaa ggtaggcata 12420
ttgtcaaaac tcagctgagt atattttgtg aatctatttc ttattgttta ctcctccact 12480
aataccacac tgtggtgact ctagtagctg tacagtaact cttaacatca tatagggcaa 12540
ttctttccac tttattgatt tatattttca gaatggcttt agcttttctt gtcccttgcc 12600
tttccataaa aattcagaat aagcttgtaa gtgtctacaa acaaacctgc cataattttg 12660
ataagaatta aagcagaggt gtccaatctt ttggcttccc tgggccacag tggaagaaga 12720
agtgtcgtgg gccacacata aaatacacac acacacacac acacacacac acacacacac 12780
acacacacac aaatggtctg tgtatagttt tcattatata tctaccacca cagataagca 12840
aaaatgtcct tgcataataa tcctaattat gcactgcccc attcagaggg tctttcaaaa 12900
tcattgaaca ggttccaagt ttgcaatcac tgatacagaa aatgtacata tctagctaaa 12960
cttcactact tttttgatat tttttattat aaaagaaaag agaacaacat aaaactagtg 13020
gggtacttga cattgttttt gagaaactaa tccatcagta tctggcttga tggaagtagt 13080
tgcaattctc agtgagttct caaggtgctc atcagatatt ttggttctaa ttttactctt 13140
cgtgttcttc atccttgaaa atagtagctc acaaatgtaa gtgctgccaa aaagcaatga 13200
catgaacaag gtgtgattgt gaagcaaggg atatttgtca ttgggaagac aggtcttaca 13260
aaagtccagt aaagaggcaa aatcaaattt ttctataagt tgaacatcag attgcagctc 13320
taggcattcc atttcaaaat tgccaggtaa catatatatg tcgactgaaa atggagttgc 13380
aaatatacca aaatattgat gatttttca gaaatcttga aatacctgtt ttcaaattcc 13440
tgtatcaaat tgaaaagcaa ggctgcgtat ttttggctgt tcacaggacc atgtttagcc 13500
aacatgtcga aatgcataaa attgtttgcc ttaatttgag cttgccataa tttcagtttc 13560
atatggaatg ctgttatggt ttgaaacatt gtattgttaa gttggttttc aacttgaaga 13620
cacaggttta actcacttaa atgggccgtc aaacccacta aaaatgctaa atctgtaagc 13680
cagttttcat tgtcaagttc tggcaccaat tttgtttgat accataaaca gcttgatttc 13740
acatcacaaa gcataaaatc tttacatttt gccttgactt aaccatctta cttctaaaaa 13800
gtgaatgact tgctagagtc agcatccata cttttaagga attcctgaaa ctagcgatga 13860
ttcaattcct gggcccttgt gaaatttaca gccttgatga caatttgcat gacgttatct 13920
acttttaaag cttgtgcaca tggattttct tgatgtatta tgcaataata cttcatcaaa 13980
tgtgagtttt gtgtggcaac tgcatcatct attaattgta caagtccctc tcttttacct 14040
accatcgcca gggcagcatc tgtagctata tcacatatgt ttacaaagga caaagaaaat 14100
tgctttaaca tatttttcac tgcttcatat aaatctcttg atttagttgt gtcttttaat 14160
agcatggtga catttcgatt tcttcagtga cattatattc atcatcaata cctctaataa 14220
aaatagcaag ttgtgccgta tctgtagtgt cagtgccttc atccatcacc aaagcataaa 14280
attttaaatt agcagtttta ctctccaaac gtctttcaat agatttccca atttctccaa 14340
ttctcctggc tatagtctgg tgagacaaac tgattttaga aatatcagtt tctcaaggca 14400
aataatatct accacatctt ccagacattg cttaataaac tcaccatcag taaatggttt 14460
tgatttcttt gctattaaat ttgctaccac ataactaggt tttaccttac gattgagtcc 14520
gagttgtaac tttttaaaaa atctttttg ttgaaaagac agacttttt tcagttctgc 14580
tattttgtcc ttacaacaca tacacaccaa atttgtcagc acgttttgc atataatgcc 14640
tcttcaaatt gtagtctttg aaaactggca caaattccgt ggaaattaag cagagtgctt 14700
cgctatttgc ctcaacaaga aaaagtcatt tgtccacttt tcattgaaca atcttccttc 14760
atccataatt tttgtttttt agggttttct ttttaagaca ttgtggaagc cattctggaa 14820
ttaaaagcat tataatagat aagcaactat atttactttt attatggaaa ttaacagata 14880
ggaaaataga acagaaagca aggtttaata atcaaataag aatacttaca tgtcttctaa 14940
ataatattaa acacctatca tctacaaagg taggttgaaa tattattgat aattgctggg 15000
ttttacttgc caaattgcca caaacacacc taatacctga cagtgtcaat tcaactgtcc 15060
gtgattagaa gataacacac tggaagtcgc acaccaccat aaaactgaag ccacacatgc 15120
gtacaaatgg cgacagtgtc tggtgtacag cagcgctctg ccttgtccag aatacacact 15180
tgaattcttt gtcacaattc acttcacgtg gcactgcaat agcgtcctct cgctctttgt 15240
tagttaattt taatggcttt taatttcttc ttgctgaact gtttgcaatt ataatgcaaa 15300
ttatggatac tagtccatta tttgtggatg tgacatactc tgattacccc tttccattcc 15360
attgttgtct acgaagttca cacttgagaa tcacatagtc aaattacaaa attacaaaaa 15420
aaattgcaaa aaaactcaaa atgttttaag aaagtttcca catttgtatt gggatacatt 15480
caaagccatc ctggactgca tgaggcctgc aggccacaag ttggacaagc ttgaattaaa 15540
ccaatagaac aatttgggta taatctatat ctttactatg ttcagccttt catcccgtga 15600

atatagtatg cctctccatt tctttagctt ttattacttt cctcaacatt ttatagtttt 15660
cagcatagag gtcctgtaca tcttttgtta gatttacacc agaaatattt catttttgtt 15720
ggagtaactg taaatgatac tgtttttctt gtattttcag atattgatta ttgttacata 15780
gaaatgtgaa taattttgtt tgttgatctt gtatcctata gccttgcaga acttacctat 15840
tcgttctaga aatttttttg tatattcctt gacattttat acattgacaa ttatgtcacc 15900
tgaaaataga gacaattcta ttatttcctt tccaatctgt atgcctttta tttctttttc 15960
ttgtctagtg tattaagaca tcaggtatgc tctttagtaa gaatgttgag agtgggcatt 16020
ttttagttct tcttgatctt ggaaaaacca ttcagtcctt catcattaaa tgtgatttaa 16080
ctgaatgatt tttttacaga ttgtctttat caaatgaagg aactgtctct ctcttcctag 16140
tttattgaga ttttatcatg acagctggaa gtacacattt taaaacaaaa catagttgtg 16200
gaagataaga gaaagttcca agcatgctgg cttgatagtc cagccccaag ttgggaaaag 16260
taattatccc tttctttttc cttctattta tggaataaaa aattaagaga aaagaatttt 16320
caaggaaatt gcattattcc ttcaaaacag gtttctagtc tttaagtatt acctactttt 16380
caaaaaaaaa tcaccacatc atggcatccc tttttcaagt tgcccatgct gtaggtgtat 16440
taaagacaga gctggtctga ggcaacatac agtctgccca tctgtcacca atccttttct 16500
actctgcaca ctcctgggga agggctaggt cttgttcctg tctattccac tggaagaaca 16560
gttccctacc acgtggagca tttgcaatta aaaggagact gagatataga ggcaggagac 16620
cacaccagat ggctgggtct ccccactccc acccccgccc cacatacact cagaagaggc 16680
taggcatcta ggatctccat tgagcatctt gaatatggct tgccataata tcatatacag 16740
tcaataaata tttgttaaat aaggatgcct cttcaatata ttttgtgcaa ccatgaagat 16800
caccacaact aatgtgagaa aaaatgttty tgttgaactc tagtctttag gcccagtggg 16860
atttatgaaa agtgccatct ctttagctga ggatgaagaa tggaagagaa tacggtcatt 16920
gctgtctcca accttcacca gcggaaaact caaggaggta tgaaaataag atgagtctta 16980
attagaaatg taaagaatga atctggggac aggtagaaag taagatcaca gtccgtttcc 17040
aaggggtagt ccactgagtt cgagcttcct aaaaatggtc ttttatcttt atgtacagaa 17100
aagacatcac aaaattcatt acaaaatgtc acttactgct ccatgctgga gaaagccata 17160
tccttctggg acttgagtct gcacatttaa ctacaggtac tgatctgttt tgtgcttaga 17220
tgttccccat cattgcccag tatggagatg tattggtgag aaacttgagg cgggaagcag 17280
agaaaggcaa gcctgtcacc ttgaaagagt aagtaggagc acagccatgg ggttctgagc 17340
tgtcatgagc ccttccagct gcctgccatg gagtcgacag tcrcactgtt gggttactcc 17400
agtgaccaga caaaagcagg gcagcgctgc aactccaaag agccacctaa gagggagtgg 17460
ctcccatgag gcggcaagtc agcaagggaa aagggccttc tctcctgtgc acaggagcca 17520
ggatttactt atctgttaac ttgtcaccat aaatattctg ggagattaaa tacatacttt 17580
agaaattaaa aaaacatgat tgtatcaaag ttttgagtgt agtggatatg gaactgtggg 17640
taagcaagca tttggtactt gttgccttgc attgggtaag atgggaaagt tacaatgggg 17700
aacttggaac aatttcaatc ccttcatggt ttttctgaga atatcagcaa actatgaact 17760
attaaacctt cccactactt ccttttcctc caatctcaaa aaagaaaggg tgctagaaat 17820
gctatgtgta gagcaagcct attatttgct gtctacaatg gtatgtgctt caattatgca 17880
ggaacgacag gtgtaatctg agcctgtcct gttcagactt gggacatgtg gtcactcagt 17940
tttgggttct ccaaatcaat gttggagaga tctatttttt ttaaccagaa cattcttgat 18000
tgtcacatct tacaaaaatg actctgctct cagcgcaact tcaggtcaga ggagctgggg 18060
atagtggggt tttccagagc attagcaggg agtgtagaga ataaaggatg atatttctag 18120
gaactcagaa cagggtgtta ctgttttgta aagtgttgaa gaggaattgg ctctgggcat 18180
agagtctgta gtcagacaac gccacctttc ttgaatccac taggaagagt taattattct 18240
actcttgttc tgctgaagca cagagcttac atatcttata tcatccacac tcaacacatg 18300
ctactgtagt tgtctgataa tgggtctctg tcttcctatg actgggctcc ttgacctcag 18360
aggtgagtct aactcagctt ggtgtctcca tcacccccag catagggcca gctccatcac 18420
tggcaccaga taaccacctt ctgagggagt agatggaaga tgattcagca gatagttctg 18480
aaagtctgtg gctctttatg tgtcttgact ggatatgtgg gtttcttgct gcatgtatag 18540
tggaaggacg gtaagaggtg ctgattttaa ttttccatat ctttctccac tcagcatctt 18600
tggggcctac agcatggatg tgattactgg cacatcattt ggagtgaaca tcgactctct 18660
caacaatcca caagacccct ttgtggagag cactaaraag ttcctaaaat ttggtttctt 18720
agatccrtta tttctctcaa taagtatgtg ggctattatt tctttctctc tttttaaaaa 18780
taactgyttt cttgacatat aattcacata tcgtataatt catccactta aaaggtacaa 18840
ttccattgtt tttaagataa tcaaaaatat gtatgaccat tactattgta aactaaaatg 18900
tttttgtcaa tctagagccc tcacacactt tagctgtcaa caccccacca caaaccccac 18960
tgccctaagc atccaataat caactttctg cctctataga tttgcctatt ctggacactt 19020
catagaaata atatcattga tttttctctg ttgtttttta ttctctattt catgagttta 19080
ttttagtctg ttatttttctt tcttttgctg gctttaggtt tcatttgctc ttcttctttt 19140
agtgttttgt ggtgtaaata attataatca atttgagata ttttcttctt ttaaatttag 19200
atattacagc tataaatttc cctctgagca ctggtttggc tacatcctgt gttttggtac 19260
atcatgcctt ctttttgttc atctcaaaac aatttcttgt tgccctttg atttctgctt 19320
tgactcactg gtcacttaaa actgtattgt ttaacttcca caaatgtatg agtttcccaa 19380

atttctttcc cttattgatt tctagtttta ttccatggaa gttgatgtac atatgctgtg 19440
ttaattctat cttgactatc atttcctgaa cagcatgatt aagttaagca gcagattatg 19500
gtctacatta atccaaaaac tctagtccaa tagataaagg ctaagaggtc agggaattta 19560
attctattac tttggtcact ccaaagactc agaaggtgcc attgatctca ctgctgtagt 19620
ggtgtttcct atgtatagac ctgcccttgc tcagtcgccg gcctgaaaga agggcaaaca 19680
tgataaaagg aatgggttcc agttgagaat catgatgttc ttattcttat tactggtaga 19740
gaaaattata attgytccag gtaaagtttg cattttcaat gatttccttt tgtttgtttt 19800
gttttyccca cagtactctt tccattcctt accccagttt ttgaagcatt aaatgtctct 19860
ctgtttccaa aagataccat aaatttttta agtaaatctg taaacagaat gaagaaaagt 19920
cgcctcaacg acaaacaaaa ggtaaaatct gatggtggtt aaatgacgat gtttaggttt 19980
tgataaattt agattttata cacatgatag agcatgtatc tgtattttta aaaataaaga 20040
cagagaactt atgtttagaa caagmgaagc catttggtag aaataaagaa ggagattggg 20100
gaaggagatg agaatgagtc agagagatag catttaaaac ttgaaatcag gcacaacaat 20160
tagtatgtca tgatataaac agtattgaga taaaatttta ccacttctct tccctttaat 20220
aaattgtcaa aggataaagt ttcctgtttg aaaatatatt ttactggtat tgtgctttcc 20280
tcatatcaca gattggtaaa gaatcatttt aagtccaaga ctcttatttt acatattctg 20340
caattaaagg tcctatgagg ctacctgccg actgctgaca tgtagtgtgt ggtaaatgtg 20400
agtgtttcac agcctggagt gaacaggggt cttctctgag aattgaggtt gcaaggctgg 20460
ctaactcagc tttgccttca cgagccctag aggccagccg aaggatgtct gcaggtcagg 20520
gagacaggac caggtaaccc agctgtcact gaagattata tagagtttga gaatgttgga 20580
atatttgaaa atgctccccc aaaaaagctg ctgatgagtt ctggaaatgt caggagatta 20640
atctatacgg acactgctga agaaaaaggt agaagaataa aagatccagt acttcttcct 20700
gggtaagcag ttatgaccag agatggaacc ggcaactctt tggccagaaa gctgtatcca 20760
aaagacagag aagatgagaa acagggaggg caaaggcgaa aaagcaattg gacatgatag 20820
ctagatttgt ttcaggaaaa catcctgctt tccaaggatt tagatgaatg tttttgttca 20880
ctggtgactc aggtaacacg tcttcaagaa gccataggga ggttgaggga gggaagtcaa 20940
gaagggaggt tgaggactgc acttttgatt tacttctgac ttcacgagtc actttctgcc 21000
aaagaaatct ctccttttgc ttctagcacc gactagattt ccttcagctg atgattgact 21060
cccagaattc gaaagaaact gagtcccaca aaggtaacca aggagtgctt ctgagggcta 21120
ctggcgggga cactaagagg gagggccttg ttctgaaaat gtgcaggaak tattccagga 21180
agatgagaat ttttgccaca tagcagaaca acacacattt agatgttata aatggtagct 21240
ggaggcactt tccagaagcc cacaggtata gccatgttcc aggctgaaag ggcaacccta 21300
agcaaaccta gaatgcttgg aggacagtca gtggtttgtg gatcacctac atgagatcaa 21360
atgccagttc tcagcctcct ccagatccac caagtgagaa cctctacttg gaaatttata 21420
tcaaacatac cgatcaggaa gcacactatc ccagtaaggg tgattttaac tggcagtact 21480
tgaaagtgtg ttcgcaaggt taatctactg caaagtttta tttttccctt tgaaatgcat 21540
aagtaactaa tgggggacac ctctgatacc atgtaaatct acttcaatct tcagtcttgt 21600
atctactagt tttatgaccc atggatggtt ttaaccaaaa ccattattac taagacagtg 21660
gcaaaatgat aaccatggtc aatttcaagc taccaagatt tggcaaccat ctcacaaaat 21720
ttttgaatat ttaacaattg gttctagaga gcaggactca gcagactcca gtataccact 21780
ttaaacatgt ccatgtctac atctacttct gtctgtctat ctatctgtca atcatctatc 21840
tgcctataat ttatcaatta atcatctatc tatctcaaca aaacttgctg tgataaagaa 21900
aatagtctat catttoactg tttcatatag aaatcactag acacatatgg ctattgagta 21960
ctggacatgt ggccaatgcc actgaagaac aatttttaag agtatttatt tttaattgaa 22020
taaaatttga atttaaatag ccacatgtgg atagtggcta ccagattgga cagcagagct 22080
cccaacttta aaattacagt tcaatttcaa ctcagtataa tggggttcaa tgtaactgag 22140
taaaataatt ggatggttga atttacccac agcagcatac agaaatattc actgataaat 22200
cagaactctg tagacctttc tcacactcat tttatattgt gtttggttgt gagttacatg 22260
attgctgcag gcaccatatt tatttctgtg ctccaggtct ctaaaggtcc taatccagtc 22320
ctgaccaaac agactagtga tggaccatcg tgagcttctc tcaggagaaa tatcaagagg 22380
gaggccaacc tgtaatcata agaacttctg ctattttaat gccattcatc agactacagt 22440
caatcaccat gcttctggct ttttgtctat ctctgctgtc ttgtacatcc tgagatagtc 22500
cattctgaga actgtaccct agatcttgta ttgcctgatg cctgtcaaag atgtaatcca 22560
tgctgcttaa gtgaggttgt gcacacaaat caccatatct cctgcaagtt tggattttga 22620
ttcagtagtt cgatggtggg gtttgagatt ctgcatttct aataagctcc cagatgtggc 22680
tggtgctgct ggtccatgaa acacactttg agtagcaaga ggtgatctgt agctcagtat 22740
tggtccttta agttccctca aacatatata gagaaaaggt cctaaatatt gcaaattctc 22800
tcaaagtttg tcaagctata ttggaattct ctcaaagtct gtcaagctct attgtagaaa 22860
atcaaatttt tattgggaaa aagcctaccc catatttact tacagataaa gtacttttag 22920
gatcattcaa ggcacacacc cataacactg agtatgtaag acagaaatgc tctctctgga 22980
aattacagca gtgctggtgc tgggatgcca tgatgaggag tgtgtggccc acaatcatgt 23040
agaccttggg aaaacctgga ttaaaatgat tttgcgtcat cctggccctg tataagatac 23100
atatcagaat gaaaaccact cccagtgtga ctttgaattg cttttccatt ttttcttctt 23160

gggattagag agcttcactt agatttcatc taagctgtga tgttgtacgt tgacctgatt 23220
tacctaaaat gtctttcctc tcctttcagc tctgtctgat ctggagctcg cagcccagtc 23280
aataatcttc atttttgctg gctatgaaac caccagcagt gttctttcct tcactttata 23340
tgaactggcc actcaccctg atgtccagca gaaactgcaa aaggagattg atgcagtttt 23400
gcccaataag gtgaggggat gacccctgga gatgaaggga agaggtgaag ccttagcaaa 23460
aatgcctcct caccactccc caggagaatt tttataaaaa gcataatcac tgattccttc 23520
actgacataa tgtaggaagc ctctgaggag aaaaacaaag ggagaaacat agagaacggt 23580
tgctactggc agaagcataa gatctttgta caatattgct ggccctggtt cacctgttta 23640
ctgttatcac aataatgcta agtaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaggagtg 23700
tggcgagaag atggccaaac aggaacagct ccagtctaca gctcccagcg tgagcaacac 23760
agaagacgaa tgatttctgc atttccaact gaggtaccgg gtgcatctca atggggattg 23820
ttggagagtg ggtgcaggac agtgggtgca gtgcacccag cctgagccaa agcagggcga 23880
ggcatcacct cacctgggaa gtgcaagggg tcagggaatt ccctttccta ggggtgacgg 23940
acagcacctg gaaaatcagg tcactcccac cctaatactg cgcttttctg atggtcttag 24000
caaacggcac accaggagat tatatcccgc gcatggctcg gagggtccta cgcccatgga 24060
gcctcgctca ttgctagcac agcagtctga gatcgaactg caaggcagca gcaaggctgg 24120
gggaggggcg cccgccattg ctaaggcttg agtaggtaaa caaagctgcc aggaagctca 24180
aactgggtga agcccaccgc agctcaagga ggtctgcctg cctctgtaga ctccacctct 24240
aggggcagag catagccaac caaaaggcag cagaaacctc tgcagactta aatgtccctg 24300
tctgacagct ttgaagagag tagtggttct cccagcacac agctggagat ctgagaacag 24360
acagactgcc tcctcaagtg ggtccctgac ccccgagcag cctaactggg aggcaccccc 24420
cagtaggggc agactgacac ctcacacggc cgggtactcc tctgagacaa aacttccaga 24480
ggaatgatca ggcagcagca tttgcgggtc accaataccg ctgttctgca gcctccactc 24540
ctgataccca ggcaaacagg gtctggagtg gacctccggc aaactccaac agacctgcag 24600
ctgaggatcc tgactgtcag aaggaaaact aacaaacaga aaggacatcc acaccaaaac 24660
ccatctgtac atcaccatca tcaaagatca aaggtagata aaaacacaaa gatgggggaa 24720
aaacagcaga aaaactgaaa aatctaaaaa tcagagcacc tctcctcctc caaaggaacg 24780
cagctccgca ccagcaacgg aaagctggat ggagaatgac tttgacgagt tgagagaaga 24840
aggcttcaga cgatcaaact actccgagct aaaggaggaa gttcgaaccc atggcaaaga 24900
agttaaaaac cttgaaaaaa gattagacaa atggctaact agaataatca atgcagagaa 24960
gtccttaaag gacctgatgg agctgaagac catggcacga gaactacgtg atgaatgcac 25020
aagcctcagt agccaattca atcaactgga agaaagggta tcagtgatgg aagatcaaat 25080
gaatgaaatg aagaaagaag agaagtttag aagaaaaaga ataaaaagaa aggaacaaag 25140
cctccaagaa atatgggact atgtgaaaag accaaatcta cgtctgattg gtgtacctga 25200
aagtgacggg gagaatagaa cgaagttgga aaacactctg caggatatta tccaggagaa 25260
cttccccaat ctagcaaggc aggccaacat tcaaattcag gaaatacaga gaacgccaca 25320
aagatactcc tcgagaagag caactccaag acacataatt gtcagattca ccaaagttga 25380
aatgaaggaa aaaatgttaa gggcagccag agagaaaggt cgggttaccc acaaacacaa 25440
acccatcaga ctaacagtgg atctctcggc agaaactcta caagccagta gagagtgggg 25500
gccaatattc aacattctta aagaaaagaa ttttcaaccc agaatttcat ttccagccaa 25560
actaagcttc ataagtgaag gagaaataaa atactttaca gacaagcaaa tgctgagaga 25620
ttttgtcacc accaggcctg ccctaaaaga gctcttgaag gaagcactaa acatggaaag 25680
gaacaactgg taccagccac tgcaaaaaca tgccaaattg taaagaccat cgaggctaag 25740
gagaaactgc atcaactaac gagcaaaata atcagctaac atcataatga caggatcaaa 25800
ttcacatata aaaatattaa ccttaaatgt aaacgggcta aatgctccaa ttaaaagaca 25860
cagactggca aactggatag agtcaagacc catcggtgtg ctgtattcag gaaacccatc 25920
tcacgtgcaa agtaacacat aggctcaaaa taaagggatg gaggaagatc taccaagcaa 25980
atggacaaca aaaaaaggca ggggttgcaa tcctactctc tgataaaaca ggctttaaac 26040
caacaaagat caaaagagac aaagaaggcc attacataat ggtaaaggga tcaattcaac 26100
aagaagagct aactatccta aatatatatg cacccaatac aggagcaccc agattcatga 26160
agcaagtctt tagagactta caaagagagt tagactccca cacaataata atggaagact 26220
ttaacaccac actgtcaaca ctagacagat caacaggaca gaaagttaag aaggatatcc 26280
aggaattgaa ctcagctctg cacaaagtgg acataataga catctacaga actctccacc 26340
ccaaatcaac agaatataca ttcttttcag caccacacca cacctattcc aaaattaacc 26400
acatagttgg aagtaaagca ctcctcagca aatgtaaaag aacagacatt ataacaaact 26460
gtctctcaga ccacagtgca atcaaactag aactcaggat tcagaaactc actcaaaacc 26520
gctcaactac atggaaactg aacaacctgc tcctgaatga ctactgggta cataacgaaa 26580
tgaaggcaga aataaagatg ttctttgaaa ccaacaagaa caaagacaca acataccaga 26640
atctctgggc cacattcaaa gcaatgtgta gagggaaatt tatagcacta aatgcctaca 26700
agagaaagca ggaaagatct aacattgaca ccctaacatc acaatgaaaa gaactagaga 26760
agcaggagca aacacattca aaagatagca gaaggcaaga aataactaag atcagagcag 26820
aactgaagga aacagagaca caaaaaaacc cttcaaaaaa atcaatgaat ccaggagctg 26880
gtttttttgaa aagatcaaca aaattgatag aatgctagca agactaataa agaagaaaag 26940

agagaagaat caaatagatg caataaaaat gataaagggg atatcaccac ccatcccaca 27000
gaaatacaaa ctaccatcag agaatactat aaacacctct atgcaaataa actagaaaat 27060
ctagaagaaa tggataaatt cctcgacaca tacactctcc caagactaaa ccaggaagaa 27120
gttgaaactc tgaatagacc aataacaggt tctgaaattg aggcaataat taatagctta 27180
ccaaccaaaa aaagtccagg accagatgga ttcaccgccg aattctacca gaggtacaag 27240
gaggacctgg taccattctt tctgaaacta ttccaatcaa tagaaaaaga gggaatcctc 27300
cctaactcat tttatgaggc cagcatcatc ctgataccaa agcctggcag agacacaacc 27360
aaaaaagaga attttagacc aatatccctg atgaacagtg atacaaaaat cctcaataaa 27420
atactggcaa accgaatcca gcagcacatc aaaaagctta tccaccatga tcaagtgggc 27480
ttcatccctg ggatgcaagg ctggttcaac atacgcaaat caataaacat aatccagcat 27540
ataaacagaa ccaacgacaa aacccacatg attatctcaa tagatgcaga aaaggccttt 27600
aacaaaattc aacagccctt catgctaaaa actctgaata aattaggtat tgatggaacc 27660
tatctcaaaa taataagagc aaatttatga caaacccaca gccaatatca tactgaatgg 27720
acaaaaactg gaatcattcc ctttgaaaac tggcacaaga cagggatgcc ctctctcacc 27780
actcctattc aacatagtgt tggaagttct ggccagggca atcaggcaag agaaagaaat 27840
aaagggtatt caattaggaa aagaggaagt caaattgtcc ctgtttgcag atgacatgat 27900
tgtatatcta gaaaacccca tcgtctcagc ccaaaatctc cttaagctga taaacaactt 27960
cagcaaagta tcaggataca aaatcaatgt gcaaaaatca caaatattct tatacaccaa 28020
taacagacaa acagagagcc aaatcatgag tgaactccca ttcacaattg cttcaaagac 28080
aataaaatac ctaggaattc aacttacaag ggatgtgaag gacctcttca aggagaatta 28140
caaaccactg ctcaatgaaa taaaagaaga tacaaacaaa tggaacaaca ttccatgctc 28200
atgggtagga agaatcaata tcatgaaaat ggccatactg cccaaggtaa tttatagatt 28260
cagtgccatc gccatcaagc taccaatgac tttcttcaca gaactggaaa aaactacttt 28320
aaagttcata tggaaccaaa aaagagcccg cattgccaag tcaatcctaa gccaaaagaa 28380
caaagccgga ggcatcatgc tacctgactt caaactatac tacaaggcta cagtaaccaa 28440
aacagcatgg tactggtacc aaaacagaga tattgatcaa tggagcagaa cagagccctg 28500
agaaagaatg ccacatatct acaaccatct gatctttgac aaacctgaca aaaacaagca 28560
gtggggaaag gattccctat ttaataaatg gtgctgggaa aactggctag ccatatatag 28620
aaagctgaaa ctggatccct tccttacacc ttatacaaaa attaattcaa gatggattaa 28680
agacttacat gttagaccta aaaccataaa aaccctagaa gaaaacctag gcaatatcat 28740
tcaatacaga ggcatgggca aggacttcat gtctaaaaca ccaaaagcaa tggcaacaaa 28800
agccaaaatt gacaaatggg atctaatgaa actaaagagc ttctgcacag caaaagaaac 28860
taccatcaga gtgaacaggc aaccgacaga atgggagaaa atttttgcaa cctactcatc 28920
tgacaaaggg ctaatatcca gaatctacaa tgatctcaaa caaatttaca agaaaaaaac 28980
acaaccccat caacaagtgg gggaaggata tgaacagaca cttctcaaaa gacatttatg 29040
cagccaatag acacatgaaa aaatgttcat catcactggc catcaaagaa atgcaaatca 29100
aaaccacaat gagataccat ctcacgccag ttagaatggc gatcattaaa aagtcaggaa 29160
acaacaggtg ctggagagga tgtggagaaa acaggaacac ttttacactg ttggtgggac 29220
tgtaaactag ttcaaccatt gtggaagtca gtgtggtgat tcctcaggga tctagaacta 29280
gaaataccat ttgacccagc catcccatta ctgggtatat acccaaagga ttataaatca 29340
tcctgctata aacacacatg cacacttatg tttattgcag cactattcac aatagcaaag 29400
acttggaacc aacccaaatg tccaataatg atagactgga ttaagaaaat gtggcacata 29460
tacaccatgg aatgctatgc agccataaaa aatgatgagt tcatgtcctt tgtagagaca 29520
tggatgaagc tggaaaccat cattctcagc aaactatggc aaggacaaaa aaccaaacac 29580
tgtatgttct cactcgtagg tgggaattga acaatgagaa cacatggaca caggaagggg 29640
aatatcacac actggggcct gttttggggt gggaggagtg gggagggata gcattaggag 29700
atataccgaa tgttaaatga cgagttaatg ggtgcagcac accaacatgg cataggtata 29760
catatgtaac aaacctgcac gttgtgtaca tgtaccctaa aacttaaagt ataaaaaaaa 29820
aaattcaaaa acctcagtgg catctaatga gaagcattta ttgctcacaa gactggatag 29880
tgagttctgc tgatactgac tggactcact ctggtctggc tatggtctga ggtagcctgg 29940
ccctgggggc gcgatggagg ctgactcagc tctccccaca cctgtctcat gttccagtca 30000
ggtagccact ggccaagaag ccaagctagg aaccagggta tctgactcct gagctaaact 30060
ctaaccctct acaatactgc ctcccaaata taacaccaag tgctaggtac atatcatcca 30120
cagttttcag acttctgccc aaactgggat tctttttagt gtgaagagac ctggcctgtg 30180
gggctgaccc tggtgtggct gtgaggcaga cacaaaggga catttacatc cagtcctgaa 30240
gattacagtc cagccctgaa gcaacaacta ggaaactatt ccaaaaggag gggatggggc 30300
tgagtgtggg gttctattct cttcataact ttaactagaa ctcaaattgt gtaccttggt 30360
agcatccaat cataaattta ttttgtcgta tttgtgatag aaaggaacaa gtttatccac 30420
aaatttattt atttatttat ttatttattt atttatttga gacagggtct gactctacga 30480
cccaagctgg agggcagtgg tgcaatctca gctcactgca aactctgcct cccaggctca 30540
agccatcctc ccgcctctgt ctcctgagta gctggaacta caggcacacg ccaccacacc 30600
cagctagttt ttgtattttt tgtagagatg ggttttcacc atgtttccca agctggtctc 30660
aaactcctca aaagagttac caagcaggac tctgcaacca ataatccttg tgtgaagagg 30720

atatttgctc ttttccctgt ttttctttct tggtacagat gtgtgacctc tttttgaaag 30780
gtgatagtga ctttggtgta ttttatttgg tggtaatggt catagcccca ttaatcacat 30840
ttcttcccat gagaaagaaa aaccactaca tggtcatgct aaggatttca gtccctgggg 30900
tgaggatggt cttgaatatc tcctacattc ataactcctc cacacatctc agtaggtcac 30960
tgagcacatc aatggacatg ccagttatta aaatacttca cgaatactat gatcatttac 31020
cagtatgagt tattctctgg agcttctaat acttcartag tactgcatgg actcagttga 31080
gagttaattc aaaatctcag attatccaat tctgtttctt tccttccagg caccacctac 31140
ctatgatgcc gtggtacaga tggagtacct tgacatggtg gtgaatgaaa cactcagatt 31200
attcccagtt gctattagac ttgagaggac ttgcaagaaa gatgttgaaa tcaatggggt 31260
attcattccc aaagggtcaa tggtggtgat tccaacttat gctcttcacc atgacccaaa 31320
gtactggaca gagcctgagg agttccgccc tgaaaggtac aagtctccag ggaaatggag 31380
ctcaccctga cccaggctgg ttcaagcata ttctgcctct cttaatctac atgacaatcg 31440
tgtggttgta caatcatttg cttgtaagtc tttttatcac aaaaaagtga taattatcaa 31500
actttacaaa ccacagacta gaaaaaacga aactacatcc atccacagtc ccagcacaag 31560
acaaagataa tcaattatgt ccctgtgggc atttttctac gcctatatag atttttaaaa 31620
attagaatgg tatcactttt tatttggttt gaattgctgc ttacttgatt taacaggaaa 31680
ctatccactg acctatatta ctataaatat acatatatat gtatatatat aaatatatat 31740
atatgtatat attgcatatg ccataaacca tttaaccatg atgttatttc aggtgtatag 31800
gctttttatt cctttctgtt ttttctatgc tgtgcccttt agctctctga atttaacaga 31860
aactttaaaa catgcttcca cattccattt gctttcaacg ttacttgcta tttcctctgt 31920
agtaattata agagtgcagg ctgaggtcct gagaagtcct catccctaat ggtttaagcc 31980
acttcactga agacacaaga cagcacaggt cctcctggtc ctatctgtgg ctgcagtcct 32040
gtgccagctc ccttatactc tcagtagaca tctcacacac tcctccttgg aggtgtcttg 32100
agcatgctct tctgggaatt cagggacaag gtcaggcctt aggcacagtt cgcactctgg 32160
atatagttgg tgttttccca ttactgtatt attaagcaaa atttagaatg aaatttttag 32220
ggtactggct ggtgattcag gatgcttggg atctagactt tcattagccc ctacctgcaa 32280
gtttgctgat gggaggaacc ttgtcttgtt ggtcatggtg tccctagtgc tagcatggag 32340
tctgcacata atacttgttc acagagtaag tcagagctga ccaagttctc tgttttctgg 32400
agtagaggac ttctatgttt cctgcaagct cagcacttcc acctcctgtg gctgcactaa 32460
tacgaaatca gagaccactc gctgtacttc actttgaatc actcagtcac caaaaagata 32520
gtgcttgcca tgtgtcagga acttggctag gcagggagaa attcatatga tttatataaa 32580
tccataaatc catatgattt acataaatcc ataaattcat gtgatatata cgtatatgtg 32640
tgtgtatata tatattagag aatgtttgac atatacacaa gtacatgtta ccgacaccag 32700
cctatagaat agttttcgtg catctccata tatctatcac tggttccaac agccatcaat 32760
ccatgttagc tgccccatcc aaatgccacc atcaccctcc tcctgactat catgttattt 32820
tgaagcaata gcctgtaaat atttcagaat gctctccaaa atataaagac tcctgtaaaa 32880
acatatgaca acaatgccat tattactttc tttgaatcaa catttttcc ttaatataat 32940
caaatattta gaaatcaaat ttgaataaaa catgggtcaa tcttcaaaga atttatagct 33000
taatggaaca gatcaaggaa agcagggatg acactacagt agggtagcat catatgccca 33060
tgtaacttat gtgacttaaa ctatcctgta agggtgtggg ggagaaagag aggaagagat 33120
ggagagaaga aaaaggaaga gaaggaggag gagaaggagg cagaggagaa ggtggacggg 33180
gaaggtagag aggaggagga ggggaattag aaaaaaagag atgacaggag aaggaaaggg 33240
aaaaataaca acttgaaata gcacaagacg ttttctcctt ctcctttctc aatgagcatg 33300
tgaccaacac aagtgtgagt tgaggcagga atccactttt ccatccatca gtcttatcat 33360
ttatgtgcct tttatagtgt gaacacatca ccaccctgaa tataatttta gtgtttagag 33420
ataaatatta tttgcaacaa tattcatctc atctcaagaa acgctcctat agggtatgga 33480
gaatttaaag gacctgtagg ttatgatgat tataacgaaa taaccaaagc aggatttcaa 33540
tgaccagccc acaaaagtat cctgtgtact actggttggg aggtggaggg gggttgttct 33600
taagtaagaa cccctaacat gtaactctgt ggtttttatr tttcattaac tatttaatct 33660
accaatatgg aactaggttc agtaagaaga aggacagcat agatccttac atatacacac 33720
cctttggaac tggacccaga aactgcattg gcatgaggtt tgctctcatg aacatgaaac 33780
ttgctctaat cagagtcctt cagaacttct ccttcaaacc ttgtaaagaa acacaggtca 33840
gtacactttc tgtatgtttt attaagaatt tttttaactg aagggtatat atttttaaa 33900
agaatatgca tgtttatctt ttaataattc attctatggg ccaaagaacc tacttggatc 33960
catctttgat cattaaggat gcttcagttc tggacttcaa aacctgtagc attaagaaca 34020
tcatgtaaag tccacacaga ttagcatgac atgattatgt gtagtctctt tgaacctgag 34080
taagtttaaa ttcagtttca agtcaattgg aaagaagtgt tttgcacaat catgaagtgc 34140
aatgattacc tggctgtgac ttaaatggtg ttctccatca ccagaacctg cagaagctct 34200
ctcatgacag tggttctcaa ccactagctg tatattggaa tcaccaggga gcttcaaaaa 34260
ttcatgatgc ctgtgacatc tcagaaattc taaactaatt aacccagagc gtgactaggt 34320
tctgtcatgc tgtcgggtga acccctgatt agttctcacg tgaagccaag gtggagaatg 34380
actaatttca ggcatttctg gtggatatga aggactacca tagagcaggg ctatccttac 34440
tccttgacct tatgttccag gtgatacatt taaagaaaga tttagaatct tttctctgaa 34500

gaagttaaag aacagatgtc attgattcat attaagcaat agcctataag tcttatttcc 34560
aggaccggtg tatttaatat gcaactctac cccttaagta cactttgtgc ttgggagagg 34620
aggaggatgg agatggttgc catcttatct atggcttcag ggcagctgtg tagctttcct 34680
atgtgtgtat tcaggcaggg ggctcagccc tgagagaaag tgggcctctg gcacacctgg 34740
gacagggaag atattccctg gcaagctctc aggcatctca ggctggcact tctttgtatc 34800
catggcaatt tgctttcccc tcactgaact gagatcagaa tgttactctg ttggtggctc 34860
ccccaacagt gaaggggtga ctcagtgaca atagtgctag aagtatgagt caaaacactg 34920
tacaacttga gaaattcccc gtttgcacta cgcttggaag ccaagaggag atgttaaaaa 34980
gaaaagaata attctttctg aagacatttc ccatcattgc acttgatggg ttcaactggg 35040
aagggttact agactctgga agttgaaaac tgcccacata attaaactgt acaacagcta 35100
ctcaggatta ccttgcaagt tttaacctat aaaaatttaa ctttatatag cacttccaaa 35160
atagtttgcc ataataccta ctaatctgga tttaattttt aaaactcatc cttttaactt 35220
aagatttaaa taaaaaaaaa aaaacacgag tccacaagaa tttgtctcag gcctggcaca 35280
gagtcagtgc tccataaata ttttgttaaa cgatggatgg tgagtgcttt tactatccag 35340
tatttaccca gcttatagat taagtatgaa gagttcaaga tacatggtgt taagagtcgt 35400
ttttatatgc ttgcaaagca tttttgtcat attttttcta ctttgcttcc atcttttctt 35460
ctttcacttc atttattaat tctccatatg cttgtttaac tattgyagat ccccttgaaa 35520
ttagacacgc aaggacttct tcaaccagaa aaacccattg ttctaaaggt ggattcaaga 35580
gatggaaccc taagtggaga atgagttatt ctaaggaytt ctactttggt cttcaagaaa 35640
gctgtgcccc agaacaccag agatttcaac ttagtcaata aaaccttgaa ataaagatgg 35700
gcttaatcta atgtactgca tgagtagttg gtgattttgt acattcattg agctctccca 35760
gagtctgtgt agagtgttgt gcattatgta gtataaagga ggtgaccagg taagtgacag 35820
ataggtagac tcagcttctc tgcttctcat aggactacct ctacccacct ctagttagca 35880
ttatcaactc ctcctgagct ctcatcagag aataaatatt tctcaacaat ttgatccata 35940
acttttaaga aaaataagaa ttatcatgat gactctaata gtgacattta tatcacgttt 36000
tatttgtaat attctataag ttttatatta agcgaagtga taaaatcccc tttacaaaaa 36060
tattatctga tgccatcctg cacactaaag agaaatctat agaactgaat gactgaaaac 36120
cagcaaataa acatttttta tcattgtaat cactgttggt gtggggcctt tgtcagaatt 36180
ccaatttgat tattaacata ggtgagagtt aatctgctgt gactttgccc attgtttgga 36240
gaaaatattc atagtttcat tctgccttct ttgaagaaca tatttttgt aacactcaac 36300
gaagcactta tcatattatt agttatgatt tattattttt accacatctc ccctgacatt 36360
tctggaacac aggaaacatg ttttcttata cgtcttgcat tccatcttca cctcccaatt 36420
gtcttaatgc aatgaacact gaataaaaaa ttgtcaattc gtcagttgat tgggcagcat 36480
gtctaaaagc actattcatt ttccttttt attctttcat tttccctcct tttctgaata 36540
ctaaagccat taggtgggtt gcagccatgt ggtagccaca cattaaggtg gacaagagag 36600
tcatggtggc tccaagtcag attccaagtg tgctggggaa ggcatccaca tggaggggca 36660
gcctgacctg gaagcgggag cccaagcaat cagagaaggg gtccacacag aggtgtggcc 36720
ttcaagagca gccagagcct aaatagggcc tggagaaccc acgtgaggtg aggagggtat 36780
ccctgagtgg gaagggatgg gtgagagttg gctacataga agggattgat cacataagta 36840
aataaagtat actggaagct aggtgtgtca cttttgcaga aaagagtcat agattcagaa 36900
ag 36902

<210> 2
<211> 1509
<212> DNA
<213> Homo sapiens

<400> 2
atggacctca tcccaaattt ggcggtggaa acctggcttc tcctggctgt cagcctggtg 60
ctcctctatc tatatgggac ccgtacacat ggacttttta agagactggg aattccaggg 120
cccacacctc tgcctttgtt gggaaatgtt ttgtcctatc gtcagggtct ctggaaattt 180
gacacagagt gctataaaaa gtatggaaaa atgtggggaa cgtatgaagg tcaactccct 240
gtgctggcca tcacagatcc cgacgtgatc agaacagtgc tagtgaaaga atgttattct 300
gtcttcacaa atcgaaggtc tttaggccca gtgggattta tgaaaagtgc catctcttta 360
gctgaggatg aagaatggaa gagaatacgg tcattgctgt ctccaacctt caccagcgga 420
aaactcaagg agatgttccc catcattgcc cagtatggag atgtattggt gagaaacttg 480
aggcgggaag cagagaaagg caagcctgtc accttgaaag acatctttgg ggcctacagc 540
atggatgtga ttactggcac atcatttgga gtgaacatcg actctctcaa caatccacaa 600
gacccctttg tggagagcac taagaagttc ctaaaatttg gtttcttaga tccattattt 660
ctctcaataa tactctttcc attccttacc ccagtttttg aagcattaaa tgtctctctg 720
tttccaaaag ataccataaa ttttttaagt aaatctgtaa acagaatgaa gaaaagtcgc 780
ctcaacgaca aacaaaagca ccgactagat ttccttcagc tgatgattga ctcccagaat 840
tcgaaagaaa ctgagtccca caaagctctg tctgatctgg agctcgcagc ccagtcaata 900

atcttcattt ttgctggcta tgaaaccacc agcagtgttc tttccttcac tttatatgaa 960
ctggccactc accctgatgt ccagcagaaa ctgcaaaagg agattgatgc agttttgccc 1020
aataaggcac cacctaccta tgatgccgtg gtacagatgg agtaccttga catggtggtg 1080
aatgaaacac tcagattatt cccagttgct attagacttg agaggacttg caagaaagat 1140
gttgaaatca atggggtatt cattcccaaa gggtcaatgg tggtgattcc aacttatgct 1200
cttcaccatg acccaaagta ctggacagag cctgaggagt tccgccctga aaggttcagt 1260
aagaagaagg acagcataga tccttacata tacacaccct ttggaactgg acccagaaac 1320
tgcattggca tgaggtttgc tctcatgaac atgaaacttg ctctaatcag agtccttcag 1380
aacttctcct tcaaaccttg taaagaaaca cagatcccct tgaaattaga cacgcaagga 1440
cttcttcaac cagaaaaacc cattgttcta aaggtggatt caagagatgg aaccctaagt 1500
ggagaatga 1509

<210> 3
<211> 502
<212> PRT
<213> Homo sapiens

<400> 3

Met Asp Leu Ile Pro Asn Leu Ala Val Glu Thr Trp Leu Leu Leu Ala
1 5 10 15

Val Ser Leu Val Leu Leu Tyr Leu Tyr Gly Thr Arg Thr His Gly Leu
20 25 30

Phe Lys Arg Leu Gly Ile Pro Gly Pro Thr Pro Leu Pro Leu Leu Gly
35 40 45

Asn Val Leu Ser Tyr Arg Gln Gly Leu Trp Lys Phe Asp Thr Glu Cys
50 55 60

Tyr Lys Lys Tyr Gly Lys Met Trp Gly Thr Tyr Glu Gly Gln Leu Pro
65 70 75 80

Val Leu Ala Ile Thr Asp Pro Asp Val Ile Arg Thr Val Leu Val Lys
85 90 95

Glu Cys Tyr Ser Val Phe Thr Asn Arg Arg Ser Leu Gly Pro Val Gly
100 105 110

Phe Met Lys Ser Ala Ile Ser Leu Ala Glu Asp Glu Glu Trp Lys Arg
115 120 125

Ile Arg Ser Leu Leu Ser Pro Thr Phe Thr Ser Gly Lys Leu Lys Glu
130 135 140

Met Phe Pro Ile Ile Ala Gln Tyr Gly Asp Val Leu Val Arg Asn Leu
145 150 155 160

Arg Arg Glu Ala Glu Lys Gly Lys Pro Val Thr Leu Lys Asp Ile Phe
165 170 175

Gly Ala Tyr Ser Met Asp Val Ile Thr Gly Thr Ser Phe Gly Val Asn
180 185 190

Ile Asp Ser Leu Asn Asn Pro Gln Asp Pro Phe Val Glu Ser Thr Lys
195 200 205

Lys Phe Leu Lys Phe Gly Phe Leu Asp Pro Leu Phe Leu Ser Ile Ile
210 215 220

Leu Phe Pro Phe Leu Thr Pro Val Phe Glu Ala Leu Asn Val Ser Leu
225 230 235 240

Phe Pro Lys Asp Thr Ile Asn Phe Leu Ser Lys Ser Val Asn Arg Met
245 250 255

Lys Lys Ser Arg Leu Asn Asp Lys Gln Lys His Arg Leu Asp Phe Leu
260 265 270

Gln Leu Met Ile Asp Ser Gln Asn Ser Lys Glu Thr Glu Ser His Lys
275 280 285

Ala Leu Ser Asp Leu Glu Leu Ala Ala Gln Ser Ile Ile Phe Ile Phe
290 295 300

Ala Gly Tyr Glu Thr Thr Ser Ser Val Leu Ser Phe Thr Leu Tyr Glu
305 310 315 320

Leu Ala Thr His Pro Asp Val Gln Gln Lys Leu Gln Lys Glu Ile Asp
325 330 335

Ala Val Leu Pro Asn Lys Ala Pro Pro Thr Tyr Asp Ala Val Val Gln
340 345 350

Met Glu Tyr Leu Asp Met Val Val Asn Glu Thr Leu Arg Leu Phe Pro
355 360 365

Val Ala Ile Arg Leu Glu Arg Thr Cys Lys Lys Asp Val Glu Ile Asn
370 375 380

Gly Val Phe Ile Pro Lys Gly Ser Met Val Val Ile Pro Thr Tyr Ala
385 390 395 400

Leu His His Asp Pro Lys Tyr Trp Thr Glu Pro Glu Glu Phe Arg Pro
405 410 415

Glu Arg Phe Ser Lys Lys Lys Asp Ser Ile Asp Pro Tyr Ile Tyr Thr
420 425 430

Pro Phe Gly Thr Gly Pro Arg Asn Cys Ile Gly Met Arg Phe Ala Leu
435 440 445

Met Asn Met Lys Leu Ala Leu Ile Arg Val Leu Gln Asn Phe Ser Phe
450 455 460

Lys Pro Cys Lys Glu Thr Gln Ile Pro Leu Lys Leu Asp Thr Gln Gly
465 470 475 480

Leu Leu Gln Pro Glu Lys Pro Ile Val Leu Lys Val Asp Ser Arg Asp
485 490 495

Gly Thr Leu Ser Gly Glu
500

<210> 4
<211> 15
<212> DNA
<213> Homo sapiens

<400> 4
gcttgtgrgg atgga 15

<210> 5
<211> 15
<212> DNA

<213> Homo sapiens

<400> 5
ccagaacsct tggac 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapiens

<400> 6
cagttgamga aggaa 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapiens

<400> 7
tgatctayaa agtca 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
ccgtacayat ggact 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
tcttatgrtt gcaaa 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
aagaggawaa ttact 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
gcagaatmgg gctag 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
tcagctcygt tgtcc 15

<210> 13
<211> 15

<212> DNA
<213> Homo sapiens

<400> 13
tgttattmtg tcttc 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapiens

<400> 14
aatgtttytg ttgaa 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
gacagtcrca ctgtt 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapiens

<400> 16
tagatccrtt atttc 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapiens

<400> 17
ataactgytt tcttg 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapiens

<400> 18
ataattgytc caggt 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapiens

<400> 19
ttgttttycc cacag 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapiens

<400> 20
gaacaagmga agcca 15

<210> 21

<211> 15
<212> DNA
<213> Homo sapiens

<400> 21
gcaggaakta ttcca 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapiens

<400> 22
tacttcarta gtact 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapiens

<400> 23
tttttatrtt tcatt 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapiens

<400> 24
actattgyag atccc 15

<210> 25
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 25
ggtgtggctt gtgrg 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens

<400> 26
ttgaaatcca tccyc 15

<210> 27
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 27
aagaacccag aacsc 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28
cggggagtcc aagsg 15

<210> 29
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 29
agaacacagt tgamg 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapiens

<400> 30
gccactttcc ttckt 15

<210> 31
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 31
gccctctgat ctaya 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapiens

<400> 32
ggattgtgac tttrt 15

<210> 33
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 33
tgggacccgt acaya 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapiens

<400> 34
ttaaaaagtc catrt 15

<210> 35
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 35
tttgcttctt atgrt 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapiens

<400> 36
ctgatgtttg caayc 15

<210> 37
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 37
tgaaagaaga ggawa 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapiens

<400> 38
ctcccaagta attwt 15

<210> 39
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 39
ccagctgcag aatmg 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapiens

<400> 40
acttcactag cccka 15

<210> 41
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 41
gtttaatcag ctcyg 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapiens

<400> 42
gtgtggggac aacrg 15

<210> 43
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 43
aaagaatgtt attmt 15

<210> 44
<211> 15
<212> DNA
<213> Homo sapiens

<400> 44

atttgtgaag acaka 15

<210> 45
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 45
agaaaaaatg tttyt 15

<210> 46
<211> 15
<212> DNA
<213> Homo sapiens

<400> 46
ctagagttca acara 15

<210> 47
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 47
ggagtcgaca gtcrc 15

<210> 48
<211> 15
<212> DNA
<213> Homo sapiens

<400> 48
taacccaaca gtgyg 15

<210> 49
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 49
gtttcttaga tccrt 15

<210> 50
<211> 15
<212> DNA
<213> Homo sapiens

<400> 50
ttgagagaaa taayg 15

<210> 51
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 51
ttaaaaataa ctgyt 15

<210> 52
<211> 15
<212> DNA
<213> Homo sapiens

<400> 52
atatgtcaag aaarc 15

<210> 53
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 53
aaaattataa ttgyt 15

<210> 54
<211> 15
<212> DNA
<213> Homo sapiens

<400> 54
aactttacct ggarc 15

<210> 55
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 55
tttgttttgt tttyc 15

<210> 56
<211> 15
<212> DNA
<213> Homo sapiens

<400> 56
agagtactgt gggra 15

<210> 57
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 57
tgtttagaac aagmg 15

<210> 58
<211> 15
<212> DNA
<213> Homo sapiens

<400> 58
accaaatggc ttckc 15

<210> 59
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 59
aaatgtgcag gaakt 15

<210> 60
<211> 15
<212> DNA
<213> Homo sapiens

<400> 60
tcttcctgga atamt 15

<210> 61
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 61
ttctaatact tcart 15

<210> 62
<211> 15
<212> DNA
<213> Homo sapiens

<400> 62
ccatgcagta ctayt 15

<210> 63
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 63
ctgtggtttt tatrt 15

<210> 64
<211> 15
<212> DNA
<213> Homo sapiens

<400> 64
atagttaatg aaaya 15

<210> 65
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 65
tgtttaacta ttgya 15

<210> 66
<211> 15
<212> DNA
<213> Homo sapiens

<400> 66
ttcaagggga tctrc 15

<210> 67
<211> 10
<212> DNA
<213> Homo sapiens

<400> 67
gtggcttgtg 10

<210> 68
<211> 10
<212> DNA

<213> Homo sapiens

<400> 68
aaatccatcc 10

<210> 69
<211> 10
<212> DNA
<213> Homo sapiens

<400> 69
aacccagaac 10

<210> 70
<211> 10
<212> DNA
<213> Homo sapiens

<400> 70
ggagtccaag 10

<210> 71
<211> 10
<212> DNA
<213> Homo sapiens

<400> 71
acacagttga 10

<210> 72
<211> 10
<212> DNA
<213> Homo sapiens

<400> 72
actttccttc 10

<210> 73
<211> 10
<212> DNA
<213> Homo sapiens

<400> 73
ctctgatcta 10

<210> 74
<211> 10
<212> DNA
<213> Homo sapiens

<400> 74
ttgtgacttt 10

<210> 75
<211> 10
<212> DNA
<213> Homo sapiens

<400> 75
gacccgtaca 10

<210> 76
<211> 10

<212> DNA
<213> Homo sapiens

<400> 76
aaaagtccat 10

<210> 77
<211> 10
<212> DNA
<213> Homo sapiens

<400> 77
gcttcttatg 10

<210> 78
<211> 10
<212> DNA
<213> Homo sapiens

<400> 78
atgtttgcaa 10

<210> 79
<211> 10
<212> DNA
<213> Homo sapiens

<400> 79
aagaagagga 10

<210> 80
<211> 10
<212> DNA
<213> Homo sapiens

<400> 80
ccaagtaatt 10

<210> 81
<211> 10
<212> DNA
<213> Homo sapiens

<400> 81
gctgcagaat 10

<210> 82
<211> 10
<212> DNA
<213> Homo sapiens

<400> 82
tcactagccc 10

<210> 83
<211> 10
<212> DNA
<213> Homo sapiens

<400> 83
taatcagctc 10

<210> 84

<211> 10
<212> DNA
<213> Homo sapiens

<400> 84
tggggacaac 10

<210> 85
<211> 10
<212> DNA
<213> Homo sapiens

<400> 85
gaatgttatt 10

<210> 86
<211> 10
<212> DNA
<213> Homo sapiens

<400> 86
tgtgaagaca 10

<210> 87
<211> 10
<212> DNA
<213> Homo sapiens

<400> 87
aaaaatgttt 10

<210> 88
<211> 10
<212> DNA
<213> Homo sapiens

<400> 88
gagttcaaca 10

<210> 89
<211> 10
<212> DNA
<213> Homo sapiens

<400> 89
gtcgacagtc 10

<210> 90
<211> 10
<212> DNA
<213> Homo sapiens

<400> 90
cccaacagtg 10

<210> 91
<211> 10
<212> DNA
<213> Homo sapiens

<400> 91
tcttagatcc 10

<210> 92
<211> 10
<212> DNA
<213> Homo sapiens

<400> 92
agagaaataa 10

<210> 93
<211> 10
<212> DNA
<213> Homo sapiens

<400> 93
aaaataactg 10

<210> 94
<211> 10
<212> DNA
<213> Homo sapiens

<400> 94
tgtcaagaaa 10

<210> 95
<211> 10
<212> DNA
<213> Homo sapiens

<400> 95
attataattg 10

<210> 96
<211> 10
<212> DNA
<213> Homo sapiens

<400> 96
tttacctgga 10

<210> 97
<211> 10
<212> DNA
<213> Homo sapiens

<400> 97
gttttgtttt 10

<210> 98
<211> 10
<212> DNA
<213> Homo sapiens

<400> 98
gtactgtggg 10

<210> 99
<211> 10
<212> DNA
<213> Homo sapiens

<400> 99

ttagaacaag 10

<210> 100
<211> 10
<212> DNA
<213> Homo sapiens

<400> 100
aaatggcttc 10

<210> 101
<211> 10
<212> DNA
<213> Homo sapiens

<400> 101
tgtgcaggaa 10

<210> 102
<211> 10
<212> DNA
<213> Homo sapiens

<400> 102
tcctggaata 10

<210> 103
<211> 10
<212> DNA
<213> Homo sapiens

<400> 103
taatacttca 10

<210> 104
<211> 10
<212> DNA
<213> Homo sapiens

<400> 104
tgcagtacta 10

<210> 105
<211> 10
<212> DNA
<213> Homo sapiens

<400> 105
tggtttttat 10

<210> 106
<211> 10
<212> DNA
<213> Homo sapiens

<400> 106
gttaatgaaa 10

<210> 107
<211> 10
<212> DNA
<213> Homo sapiens

<400> 107
ttaactattg 10

<210> 108
<211> 10
<212> DNA
<213> Homo sapiens

<400> 108
aaggggatct 10

<210> 109
<211> 3000
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (30)..(30)
<223> PS1: polymorphic base A or G

<220>
<221> misc_feature
<222> (61)..(120)
<223> n's represent sequence between PS1 and PS2

<220>
<221> allele
<222> (150)..(150)
<223> PS2: polymorphic base C or G

<220>
<221> misc_feature
<222> (181)..(240)
<223> n's represent sequence between PS2 and PS3

<220>
<221> allele
<222> (270)..(270)
<223> PS3: polymorphic base G or A

<220>
<221> misc_feature
<222> (301)..(360)
<223> n's represent sequence between PS3 and PS4

<220>
<221> allele
<222> (390)..(390)
<223> PS4: polymorphic base C or T

<220>
<221> misc_feature
<222> (421)..(480)
<223> n's represent sequence between PS4 and PS5

<220>
<221> allele
<222> (510)..(510)
<223> PS5: polymorphic base A or C

<220>
<221> misc_feature
<222> (541)..(600)
<223> n's represent sequence between PS5 and PS6

<220>
<221> allele
<222> (630)..(630)
<223> PS6: polymorphic base T or C

<220>
<221> misc_feature
<222> (661)..(720)
<223> n's represent sequence between PS6 and PS7

<220>
<221> allele
<222> (750)..(750)
<223> PS7: polymorphic base C or T

<220>
<221> misc_feature
<222> (781)..(840)
<223> n's represent sequence between PS7 and PS8

<220>
<221> allele
<222> (870)..(870)
<223> PS8: polymorphic base G or A

<220>
<221> misc_feature
<222> (901)..(960)
<223> n's represent sequence between PS8 and PS9

<220>
<221> allele
<222> (990)..(990)
<223> PS9: polymorphic base T or A

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> n's represent sequence between PS9 and PS10

<220>
<221> allele

<222> (1110)..(1110)
<223> PS10: polymorphic base C or A

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> n's represent sequence between PS10 and PS11

<220>
<221> allele
<222> (1230)..(1230)
<223> PS11: polymorphic base C or T

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> n's represent sequence between PS11 and PS12

<220>
<221> allele
<222> (1350)..(1350)
<223> PS12: polymorphic base C or A

<220>
<221> misc_feature
<222> (1381)..(1440)
<223> n's represent sequence between PS12 and PS13

<220>
<221> allele
<222> (1470)..(1470)
<223> PS13: polymorphic base C or T

<220>
<221> misc_feature
<222> (1501)..(1560)
<223> n's represent sequence between PS13 and PS14

<220>
<221> allele
<222> (1590)..(1590)
<223> PS14: polymorphic base G or A

<220>
<221> misc_feature
<222> (1621)..(1680)
<223> n's represent sequence between PS14 and PS15

<220>
<221> allele
<222> (1710)..(1710)
<223> PS15: polymorphic base G or A

<220>
<221> misc_feature
<222> (1741)..(1800)
<223> n's represent sequence between PS15 and PS16

<220>
<221> allele
<222> (1830)..(1830)
<223> PS16: polymorphic base A or G

<220>
<221> misc_feature
<222> (1861)..(1920)
<223> n's represent sequence between PS16 and PS17

<220>
<221> allele
<222> (1950)..(1950)
<223> PS17: polymorphic base C or T

<220>
<221> misc_feature
<222> (1981)..(2040)
<223> n's represent sequence between PS17 and PS18

<220>
<221> allele
<222> (2070)..(2070)
<223> PS18: polymorphic base C or T

<220>
<221> misc_feature
<222> (2101)..(2160)
<223> n's represent sequence between PS18 and PS19

<220>
<221> allele
<222> (2190)..(2190)
<223> PS19: polymorphic base T or C

<220>
<221> misc_feature
<222> (2221)..(2280)
<223> n's represent sequence between PS19 and PS20

<220>
<221> allele
<222> (2310)..(2310)
<223> PS20: polymorphic base A or C

<220>
<221> misc_feature

<222> (2341)..(2400)
<223> n's represent sequence between PS20 and PS21

<220>
<221> allele
<222> (2430)..(2430)
<223> PS21: polymorphic base G or T

<220>
<221> misc_feature
<222> (2461)..(2520)
<223> n's represent sequence between PS21 and PS22

<220>
<221> allele
<222> (2550)..(2550)
<223> PS22: polymorphic base A or G

<220>
<221> misc_feature
<222> (2581)..(2640)
<223> n's represent sequence between PS22 and PS23

<220>
<221> allele
<222> (2670)..(2670)
<223> PS23: polymorphic base G or A

<220>
<221> misc_feature
<222> (2701)..(2760)
<223> n's represent sequence between PS23 and PS24

<220>
<221> allele
<222> (2790)..(2790)
<223> PS24: polymorphic base T or C

<220>
<221> misc_feature
<222> (2821)..(2880)
<223> n's represent sequence between PS24 and PS25

<220>
<221> allele
<222> (2910)..(2910)
<223> PS25: polymorphic base T or C

<220>
<221> misc_feature
<222> (2941)..(3000)
<223> n's represent sequence 3' to PS25

<400> 109

tgggtaaaga tgtgtaggtg tggcttgtgr ggatggattt caattattct agaatgaagg 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
cacttgagtt tctgataaga acccagaacs cttggactcc ccgataacac tgattaagct 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
gcttggctga agactgctgt gcagggcagr gaagctccag gcaaacagcc cagcaaacag 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
actgctgtgc agggcaggga agctccaggy aaacagccca gcaaacagca gcactcagct 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
taaaaggaag actcacagaa cacagttgam gaaggaaagt ggcgatggac ctcatcccaa 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
cctgagtaac tcaccagccc tctgatctay aaagtcacaa tccctgtgac ctgatttctg 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
tttcactttg tagatatggg acccgtacay atggactttt taagagactg ggaattccag 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
tgcttgagct tcctcttttg cttcttatgr ttgcaaacat cagcttagtt ccatcagtaa 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
acagagagag gttctctgaa agaagaggaw aattacttgg gagtagaata ttgcaatggg 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
cttcctgggt gtggctccag ctgcagaatm gggctagtga agtttaatca gctccgttgt 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
gaatcgggct agtgaagttt aatcagctcy gttgtcccca cacagaacgt atgaaggtca 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
cagaacagtg ctagtgaaag aatgttattm tgtcttcaca aatcgaaggg taagcatcca 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
caccacaact aatgtgagaa aaaatgttty tgttgaactc tagtctttag gcccagtggg 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
tccagctgcc tgccatggag tcgacagtcr cactgttggg ttactccagt gaccagacaa 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
ccacaagacc cctttgtgga gagcactaar aagttcctaa aatttggttt cttagatcca 1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800
aagttcctaa aatttggttt cttagatccr ttatttctct caataagtat gtgggctatt 1860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1920
atttctttct ctctttttaa aaataactgy tttcttgaca tataattcac atatcgtata 1980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040
cttattactg gtagagaaaa ttataattgy tccaggtaaa gtttgcattt tcaatgattt 2100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
caatgatttc cttttgtttg ttttgtttty cccacagtac tctttccatt ccttacccca 2220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
aaagacagag aacttatgtt tagaacaagm gaagccattt ggtagaaata aagaaggaga 2340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2400
gagggccttg ttctgaaaat gtgcaggaak tattccagga agatgagaat ttttgccaca 2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2520
agttattctc tggagcttct aatacttcar tagtactgca tggactcagt tgagagttaa 2580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2640
cccctaacat gtaactctgt ggtttttatr tttcattaac tatttaatct accaatatgg 2700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2760
taattctcca tatgcttgtt taactattgy agatcccctt gaaattagac acgcaaggac 2820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2880
cctaagtgga gaatgagtta ttctaaggay ttctactttg gtcttcaaga aagctgtgcc 2940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3000